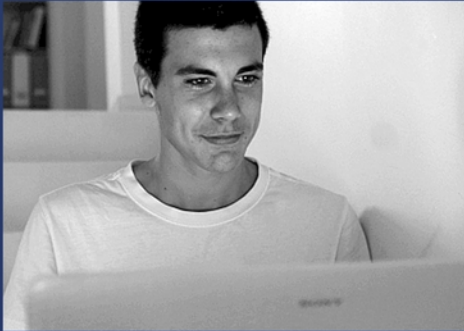


Hacker Highschool

SECURITY AWARENESS FOR TEENS



LESSON 15 DOXING

DRAFT



HACKING IS LEARNING
www.hackerhighschool.org

ISECOM

Creative Commons 3.3 Attribution-Non-Commercial-NoDerivs ISECOM

WWW.ISECOM.ORG - WWW.OSSTMM.ORG - WWW.HACKERHIGHSCHOOL.ORG - WWW.BADPEOPLEPROJECT.ORG - WWW.OSSTMMTRAINING.ORG



WARNING

The Hacker Highschool Project is a learning tool and as with any learning tool there are dangers. Some lessons, if abused, may result in physical injury. Some additional dangers may also exist where there is not enough research on possible effects of emanations from particular technologies. Students using these lessons should be supervised yet encouraged to learn, try, and do. However ISECOM cannot accept responsibility for how any information herein is abused.

The following lessons and workbooks are open and publicly available under the following terms and conditions of ISECOM:

All works in the Hacker Highschool Project are provided for non-commercial use with elementary school students, junior high school students, and high school students whether in a public institution, private institution, or a part of home-schooling. These materials may not be reproduced for sale in any form. The provision of any class, course, training, or camp with these materials for which a fee is charged is expressly forbidden without a license, including college classes, university classes, trade-school classes, summer or computer camps, and similar. To purchase a license, visit the LICENSE section of the HHS web page at <http://www.hackerhighschool.org/licensing.html>.

The Hacker Highschool Project Project is an open community effort and if you find value in this project, we ask that you support us through the purchase of a license, a donation, or sponsorship.



Table of Contents

Introduction.....	5
What Is Doxing?.....	6
Web Databases and Cached Pages.....	7
Web Cache.....	7
Unsafe Websites.....	8
HTML Attacks.....	9
Investigating Key People.....	9
Compiling E-mail Addresses.....	10
Searching for Information on the Internet.....	13
Examining Documents.....	13
P2P Networks.....	16
Sharing Documents.....	17
Fake Files.....	18
Former Employees.....	18
Password Protected Documents.....	18
Search Engine Hacking.....	19
What kind of information can you find by searching?.....	19
How Search Engines Function.....	20
Basic Operators.....	20
Advanced Operators.....	21
The Dark Side.....	21
Useful Tools.....	21
Make Web Searches Easy: Customize Them.....	21



Contributors

Marta Barceló, ISECOM

Pete Herzog, ISECOM

Bob Monroe, ISECOM

Marco Ivaldi

Aneesh Dogra

Simone Onofri

Bob Monroe

Alfonso Arjona

Dominique C. Brack

Greg Playle

ISECOM



Introduction

Doxing or **document grinding** is one of the most important tasks for a hacker. Simply having a good knowledge of tools, protocols, and hacking techniques is not enough; you must also gather information about the target.

This is doxing someone or an organization – gathering information about that entity, organizing the information, and using that information to find more. You can learn an awful lot about someone or something with a bit of digging. That might be someone you meet, someone who wants to be “friends” with you, or a potential employer. For a start, search the Web for “live tweet” “out” and “married man”. Decide for yourself how you would like that situation to apply to you.

The purpose of this lesson is to give you an understanding of some of the basic techniques used for retrieving information about a target. Knowing how this information is gathered will help you to ensure that you don't disclose sensitive data about yourself, your family or friends.

In each episode of Disney's “Phineas and Ferb,” Perry the Platypus is always caught in some trap created by the evil scientist. The traps are simple: nets, trap doors, capture chairs. Now, with proper doxing, Perry would already know what sort of traps his evil nemesis has placed.

Perry discovers a major purchase of thumbtacks and a massive electromagnet from a dummy company in China. Email traffic from the evil scientist's outbound POP servers contain lengthy discussion of thumbtack exploits with someone named The Joker. A consultant was hired two days ago to set up the electromagnet below a trapdoor.

With the help of some basic doxing, Perry the Platypus could at least try to avoid easy capture during his encounters with the crazy doctor. You can too. Just don't barge through the front doors of your target and demand their data, or you'll get a good laugh from the guys with the badges.

Sun Tzu wrote in **The Art of War**, “All warfare is based on deception.” The more you know about your target, the more likely your success will be. Research your target ahead of time or you could do time elsewhere.



What Is Doxing?

When you lose something important like your homework, one of your favorite shoes (it's always the left shoe that goes missing for some reason), your house key or anything else, your natural reaction is to (freak out then) look every place you can think of. You spend frantic minutes sifting through paperwork in your backpack, under piles of dirty laundry, under the sofa, in the cat litter box and everywhere else until you find the missing item exactly where you put it for safekeeping. Doxing is a similar process of trying to find critical information about a target (or keeping that data away from others). You can think of it as "hide and seek" in the digital world, just with a lot more (cheating) peeking.

The goal of this hide and seek game is either to hide or to seek as much information about your target as possible. This information can include:

- 1. Network details:** topology, gateways, IP addresses, access points, internet and intranet entry points, honey pots, IDS and other threat detection systems, configurations of routers and other equipment that passes data, hidden data channels, target wireless communications, VPN, RAS and APN accesses, web (IPv4 and IPv6), FTP and other Internet presence such as IRC, P2P, data banks, cloud, VOIP, and other acronyms.
- 2. Email information:** addressing scheme, internal email accounts and aliases, role accounts, accounts that don't seem to belong to anyone, email rerouting structure, email header information, messages, attachments and related policies, email clients (Mail User Agents or MUAs) in use including webmail platforms, vacation auto-reply configurations, mailing list configurations, internal email systems compared to external email, email logging files, antispam and antivirus solutions, encryption mechanisms used for communications, etc.
- 3. Target organizational structure:** who are the important folks, who are the worker bees, who are the mailroom clerks, and who are the secretaries with steely gray eyes that watch your every movement like that one mean old lady who works at the library. She's scary and I think she knows where you live. You may also want to know who holds the keys or access codes so Neo can meet the creator of the Matrix. The key holder might just be the nightly janitor crew.
- 4. Hidden data:** Shhh, this is the stuff that targets do not want the public to know about or information that you don't want others to see, like that photo of you with spaghetti coming out of your nose or the report card that "your dog ate." Everyone has secrets and targets certainly have plenty that they don't want others to know about. Depending on whether you are working for the target, working against the target, or are the target, you will want to know as much about these tasty bits of data as possible.
- 5. Public data:** Every organization wants to boast about their success and capabilities. There are plenty of places to look for pieces of helpful information within these bragging articles. Even the tiniest interview given by the new CEO in some obscure newspaper can yield critical details about your target. This topic area includes metadata, search engines, public relations, investor relations, personal employee blogs, mailing list archives, forums, wikis, public filings and digital dumpster diving into Internet databases and cache.



Web Databases and Cached Pages

You will encounter this quote several times in your career: **Knowledge is power**. And for a hacker it is the key to breaking into – or protecting – a system.

Every single day at the Internet is Harvest Time, and there are tons of sensitive information regarding key people available...if you know how and where to search. House MD has a quote that explains this perfectly: "Everybody does stupid things; it shouldn't cost them everything they want in life." And that's something that happens every day.

Imagine you are the CTO of a financial company, have a great salary, paid expenses, medical insurance, free flights all around the world and a superb car. Let's say it's the new BMW M3. This car is fantastic, but as many thing you will own in your life, it can be improved. So, you join a forum and start searching information about high performance exhaust systems, shock absorbers, superchargers, racing tires and all kind of extras for your car. After a few days, you decide to post a picture of your beauty, and once you become an expert tuning your car, it's time to help newcomers. If someone requests a body shop, you provide the name of the establishment you uses. If it's a question about a mechanical problem, you will recommend a specialist giving the name, phone number or address.

Now, everyone at the forum will think you are a nice guy. But (believe me) you are in trouble. Never think someone determined won't spend the whole day at home thinking about how to steal information. Commonly they take a look at the building where you work and at all cars in the parking lot (car park). It won't be difficult for him to see your car, and search for your license plate number on the Internet. In just a few hours, he may find information about fines, but for sure he will find those pics of your car (in front your home!), know where you used to go (because you recommended it) or if you have an alarm installed. This will help him to steal your car, or just to break a window and grab the building access card. He will even search for you in social platforms, looking for this kind of information and pictures you have shared. Why pictures? Because he will recognize you, discover your friends and search on their profiles for more information about you.

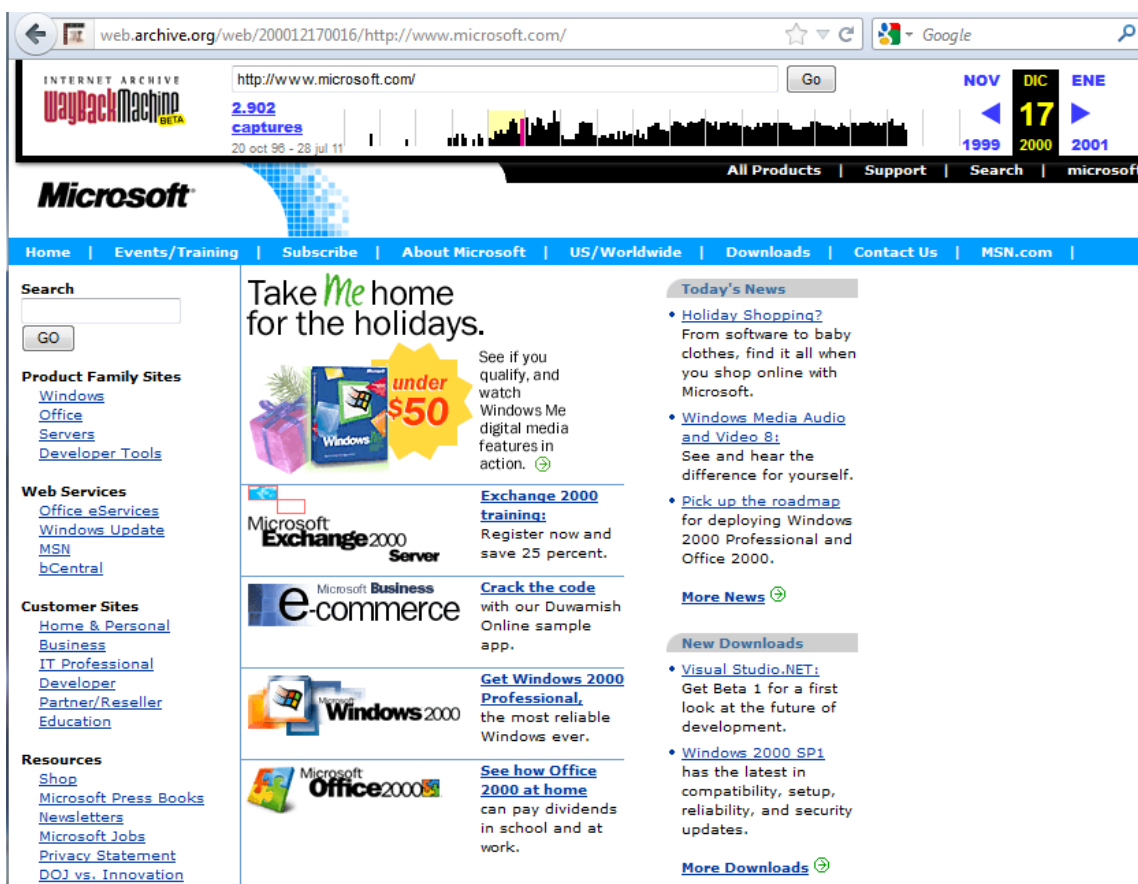
At this moment, you decide to delete all posts in forums containing sensitive information, and restrict your profile to "just for friends."

But it's too late.

Web Cache

Have you ever performed a search, found the result, and when you click the link the page shows a non-authorized message or even that the server is down? Almost every search engine has an option to show the cached version of the page. Their cache holds a copy of a page stored at the library of the search engine. It's quite useful when a server is down or too busy, but it contains old information...the kind of information you have deleted. You can request to have this cached version deleted, or wait until a new cache is created. Meantime, your sensitive information is exposed. (Don't blame the search engine. It's your fault.)

Want to see something fun? Point your browser to www.archive.org Check out the cached version of websites, even years old versions. For example, this one is Microsoft in December 2000.



But let's be clear: a cached version of a web page is not bad. It's something very useful and is not intended to expose sensitive information. If someone posted the page, it will appear. Remember that. If you put it online, it's going to stay online.

Unsafe Websites

Building a website isn't hard; you built basic pages early in this course. There are several software suites that make this an easy task, or you can create your site from scratch. All but static "brochure" sites will use a database to store information, profiles, accounts, posts and documents. The catch is that the application needs to know how to connect to the database; this information is stored in a configuration file.

This configuration file, like every file on every modern computer, must have **permissions**. Basically this means what are you allowed to do with the file. In Linux/UNIX, permissions can be Read, Write or Execute, and are set for three kind of users: Owner, Group, and Others, on Linux/UNIX. Windows generally employ user groups and Full Control/Modify/Read/Write /Special permissions for Windows. For Windows, in Windows Explorer, right-click the file or folder and select "Properties." The security tab will show the file or folder permissions. For Linux/UNIX, at the command line, issue the command

```
ls -al <filename>
```

to get the long format directory listing of all files in that directory. The left side of the listing will show the Read/Write/Execute permissions for Owner/Group/Others, respectively.

When you're browsing a site, you are a member of Others or Guests or some such menial group, and you shouldn't have permissions to access the configuration file. But if the site is misconfigured and the configuration file has Read permissions for everyone (you won't be surprised to find out how frequently this happens, will you?), anyone can read this file. The



configuration files contain the username and password for the database, IP or server name, file paths, and more.

With this information, a hacker can dump the whole database.

HTML Attacks

You've read about cross-site scripting, SQL injection and similar exploits. Generally we call these **HTML attacks**. It's not hard to understand how they work, so we'll give you another example.

The Sith council wants to have revenge. All Jedi must be exterminated so they can rule the galaxy. Darth Vader has been searching for information about Jedis and apprentices on the Galactic Internet, finds but no sensitive data. He goes to the Jedi Academy website, reads the whole site and tries to join the forum, but nothing works for him. (Don't mess with Vader, he will find your lack of faith disturbing) But the Dark Side is wise. Vader has a sense about the search box. First, he tries "Yoda," but besides a picture of the Jedi Master, he gets no interesting results. Then, he tries something different: he searches for the ' symbol (that's a single quote 0 and...gotcha! A SQL error message is returned. Vader has a look at it, and he decides to perform a new search using a specially crafted string as input parameter. Now the screen is filled with names, aliases, emails, addresses and all kinds of great stuff.

Is the ability to destroy a planet insignificant next to the power of a SQL injection?

Exercise

- 15.1 Perform a search, and select one of the results. Compare the results from cached and live versions, particularly when it comes to key people.

Investigating Key People

An often overlooked aspect of doxing is learning who the players are and who to watch out for. Think of a corporation as your school system. Your teacher is a company employee just doing their job day in and day out. These folks are fairly open with their digital social life and don't mind telling their life story on Facebook, Twitter, or their own blog. These people report to the school organizers otherwise known as a middle manager or some fancy name like "team coordinator" in the business world.

The organizers/managers report to other higher paid school organizers, known as upper management. In the world of business, upper managers usually have their own secretaries who schedule their golf tee times and weed out unwanted telephone calls and emails. Schools don't really have upper managers (thank goodness).

Above the upper managers, there is the Chief Executive Officer (CEO). Consider the CEO to be the school principal. The CEO is challenged with everyday operations and meeting the shareholders' goals. Some organizations hire and fire CEOs on a monthly basis. You might be under the impression that the CEO/Principal is all-powerful. Wrong!

These folks have to report to the Board of Directors/Superintendent. Both the Board of Directors and the Superintendent are outside the realm of reality. Don't waste your research time on anyone with this title. You will want to focus on the important people; these are the workers and the upper management. You might be wondering why these two groups are going to be your focus. The two groups share different levels of information but execute their assigned tasks in vastly different ways.



As mentioned earlier, workers are fond of sharing their lives in the digital world. There are endless web blogs, picture sharing pages, and hobby groups that love sharing that type of information. There isn't a single employee out there that isn't looking to move upwards in the work environment so locating their resume is simple. Just pose as an employer looking to hire and www.Monster.com, www.Dice.com; www.Ladders.com will start feeding you resumes like corn flakes.

Upper management is the same way, they too are always looking for a leg up in society so their resumes will be located in the higher end of pay scales. Upper managers don't waste their time on Facebook or blogs, it takes time away from their golf game. Believe it or not but many business transactions happen over 18 rounds of hitting a little white ball.

Upper management folks are the "tweens" because they use technology because they have to but don't totally understand it. You may come across one or two tech savvy business folks but that is a long shot outside of the technology field. We are looking for the breed known as **Grey Hairs**. These are the guys who can tell you everything about the company, down to an audit conducted in 1962 on the shipping department. These guys need to work, want to work, and are valuable to the organization due to their vast knowledge of the company. The Grey Hairs also know where all the bodies are buried. They use typewriters and pencils instead of keyboards but can compute complex functions before you can input the data. The one flaw of Grey Hairs is that they are forced to use computers to send in reports or emails. Grey Hairs use weak passwords, sloppy passwords, or easy to remember passwords. The Grey Hairs don't rely on a secretary unless they really have to.

Compiling E-mail Addresses

Have you seen **Spaceballs**? If not, consider it research. There's a scene that exposes a serious security problem. Dark Helmet demands the password to the "air shield" and King Roland is telling him:

Roland: One.

Dark Helmet, Colonel Sandurz: One.

Roland: Two.

Dark Helmet, Colonel Sandurz: Two.

Roland: Three.

Dark Helmet, Colonel Sandurz: Three.

Roland: Four.

Dark Helmet, Colonel Sandurz: Four.

Roland: Five.

Dark Helmet, Colonel Sandurz: Five.

Dark Helmet: So the combination is...one, two, three, four, five? That's the stupidest combination I've ever heard in my life! That's the kind of thing an idiot would have on his luggage!

(President Skroob enters the room)

President Skroob: Did it work? Where's the king?

Dark Helmet: It worked, sir. We have the combination.

President Skroob: Great. Now we can take every last breath of fresh air from Planet Druidia. What's the combination?

Colonel Sandurz: 1-2-3-4-5

President Skroob: 1-2-3-4-5?

Colonel Sandurz: Yes!

President Skroob: That's amazing. I've got the same combination on my luggage!



Well, you may think this is about passwords. That's right, but you have to think about this: If a hacker wants or needs information, all compiled information is a plus. And one of the most interesting resources is finding e-mail accounts. Why?

It's simple. If you own a list of e-mails at a target you will be able to find all the key people and employees. But wait, there's more! Think about this: What are email accounts used for?


- A site may use e-mail as the username to log in.
- If you forgot your password, you can retrieve it using your e-mail account.
- You need an e-mail address to receive pictures and jokes about cats.


Can't see the point? Consider this:


- Many people will use their business e-mail to register in Facebook, Google+, Twitter or forums.
- Also, they will provide this e-mail to receive "Amazing offers," then complain about receiving spam.
- If you need a credit card, you also have to provide an e-mail account.
- Want to buy stuff from an online shop? Provide an e-mail!
- Have to stay at the office working this night? Order a pizza! Your e-mail is your username!
- Complaining about the service? Send an e-mail.
- ...

Get it? Your email address is everywhere! Now be honest. Are you using the same password everywhere? If your answer is "yes" or "yes in most cases" you are in trouble, even if 12345 is not your password. You don't know if all information stored in those sites is safe, but if one site is hacked, somebody's got your password. Remember what happened with Sony's Playstation Network? Names, passwords, usernames, email accounts and billing information were all stored in plain text! (Duh.)

So finding email accounts is very interesting for a hacker. Here's an example: search for emails containing the "@hotmail.com" string, and the keyword phpBB.







Search About 1,080,000,000 results (0.35 seconds)

Everything
 Images
 Maps
 Videos
 News
 Shopping
 More
 Show search tools

[Dedicated Hosting Servers From Wooservers! Core i5 - \\$90.99, Dell ...](#)
[adminspt.net/.../4593-dedicated-hosting-servers-from-wooservers-c...](#)
 1 post - 1 author - 17 Feb

Msn: wooservers@**hotmail.com** ... Invision Power Services (IPB), |---- vBulletin, |---- **PHPBB**, |---- Wordpress, |---- Other, |---- SMF, |-- Hosting ...

[Need help on a drupal/zencart site. integration and social networking](#)
[drupal.org > Services > Paid Drupal services](#)
 24 May 2008 – A. Keep ZenCart + WordPress +**PHPbb**. B. Keep ZenCart ... help on theming etc. let me know. and please email me at the_az@**hotmail.com** ...

[\[احصاء\] ما هي المجلة المفضلة لديك ?](#)
[www.traidnt.net/vb/traidnt372316/](#) - Translate this page
 15 posts - 5 authors

... alraheeb.7 **hotmail.com**. Facebook ... سبحان الله ويحمده . amresam02 **hotmail.com**
 ... اقتراضي. انا التي ارها قوية ونها صعب تتذكر او تختبر هي +**phpbb**. وضاح مراد

[cherche developpeur \[php/sql\]pour mon site ??? : Projets ...](#)
[forum.phpfrance.com/.../cherche-developpeur-po...](#) - Translate this page
 4 posts - 3 authors - 1 Apr 2006

je possède déjà system chat flash +**phpbb** forum à insérer donc dans le site. ... contact

Amazing, isn't it? This is a simple search, and we've found e-mail accounts, Facebook references, public forums.

This is a simple hacker trick: search for interesting e-mails related to a site. If you don't use your e-mail carefully, this will give out your personal profile, likes, dislikes, preferences, and so on. And if a hacker finds a security hole in one site, he may be able to obtain your password.

How this will be performed? Quite simple. Let's say your name is James T. Kirk, and your are working for the United Federation of Planets. A hacker named Khan knows the domain is unitedfederationofplanets.org, but has no idea about your e-mail. Browsing the site, he will find you have been promoted to Captain and enrolled at USS Enterprise. Well, he will use a search engine to find strings like:

- James.Kirk@unitedfederationofplanets.org
- JamesTKirk@unitedfederationofplanets.org
- JTKirk@unitedfederationofplanets.org
- JimKirk@unitedfederationofplanets.org
- ...

He'll find your email in sites and forums. Remember when you cheated the Kobayashi Maru Test? It's explained in several bulletin boards! Also, your name appears at the "Most Wanted" list at the Klingon Army website, and your home page at the Starfleet Academy contains your email because students need to contact you. Remember when Khan activated the Genesis Device at Ceti Alpha VI? Nothing to do with Khan's inexperience in space combat: he found a security hole at the online shop where you usually buy Romulan Ale, and using the Reliant's computer hacked your password (the same you used at the Enterprise), reached the Weapons Control Console and activated the device.

Now, a court-martial awaits you.



Exercises

- 15.2 Perform a search on the Internet using your email address as parameter. How many results do you get?
- 15.3 Can you find information about yourself regarding your preferences, opinions and sensitive information?
- 15.4 Can you find your home address in those results?
- 15.5 In all those sites, are you using the same username or password?

Searching for Information on the Internet

This would be an easy lesson if Google could be the solution to all your research questions. Unfortunately Google Inc. is not nearly the answer to all research. Search engines look for key words within web published documents and pages. This leads to some information being omitted or flat out wrong. One trick that works really well is to find a published resource that has similar information on your target and use their bibliography or cited sources to locate even more information.

Search engines only cover the top layer of Internet available information. To get down into the weeds for target research you will need much better tools. That information will be located on special databases that can only be accessed through subscriptions. If you don't want to shell out several hundred dollars to access to those special databases, you can use your local library's online content.

You will be amazed at the information your local library has available to you online. Universities pay for subscriptions to these expensive databases for use as research tools for their students and teachers. Gale, Business Source Premier, ERIC, and JSTOR are just a small sample of databases that specialize in particular information. The difficult part is locating these repositories of information and gaining access to them.

One very valuable advantage you may have is your school email address ending with .edu. That education identifier on your personal school email address gets you free access to many places the general public can't easily reach. On the Internet, there are two main areas of information. The one you are most familiar with is the web pages indexed by Google and other search engines. You get millions of pages that are delivered to you whenever you do a search. But beneath all those pages is 500 times more data that search engines cannot show you. That information is located in the databases mentioned above. Your library isn't what it used to be, just a place to borrow books. Your library is your new best friend.

As an added bonus, librarians are trained researchers themselves. Based on the type of library, librarians can help do some of the heavy lifting research for you. Just provide the librarian with your topic with clear guidance on what you are really looking for. Within a few days, they can give you guidance to places you've never heard of. This is a good thing, trust us.

Examining Documents

Most of the files we use every day may contain sensitive information. For example, every file in your digital music library contains information about the song's name, artist, album, audio settings for your player – or even the owner.

Metadata search is one tool used to learn more about documents or files. The most important point of this topic is that “computers never forget.” Document revision history is



contained in almost every file ever created, as is ownership. Decompiling a file is a great way to search for text strings within that file; documents are the same way. Both Microsoft and the CIA have learned this lesson the hard way.

Depending on target's file format, a decent text editor will suffice for some of the heavy lifting, since metadata is viewed along with the document itself. If you really need to dig deep, forensic tools can come into play. In fact, forensic tools can help you out through this entire documentation examination process. You might find out who actually wrote particular documents. You couldn't possibly think that the company CEO wrote all those standards and visions, could you? You will want to know the brains behind the creator of documents. Besides locating the creator of certain documents, the next critical observation is to figure out who edited that document.

Exercise

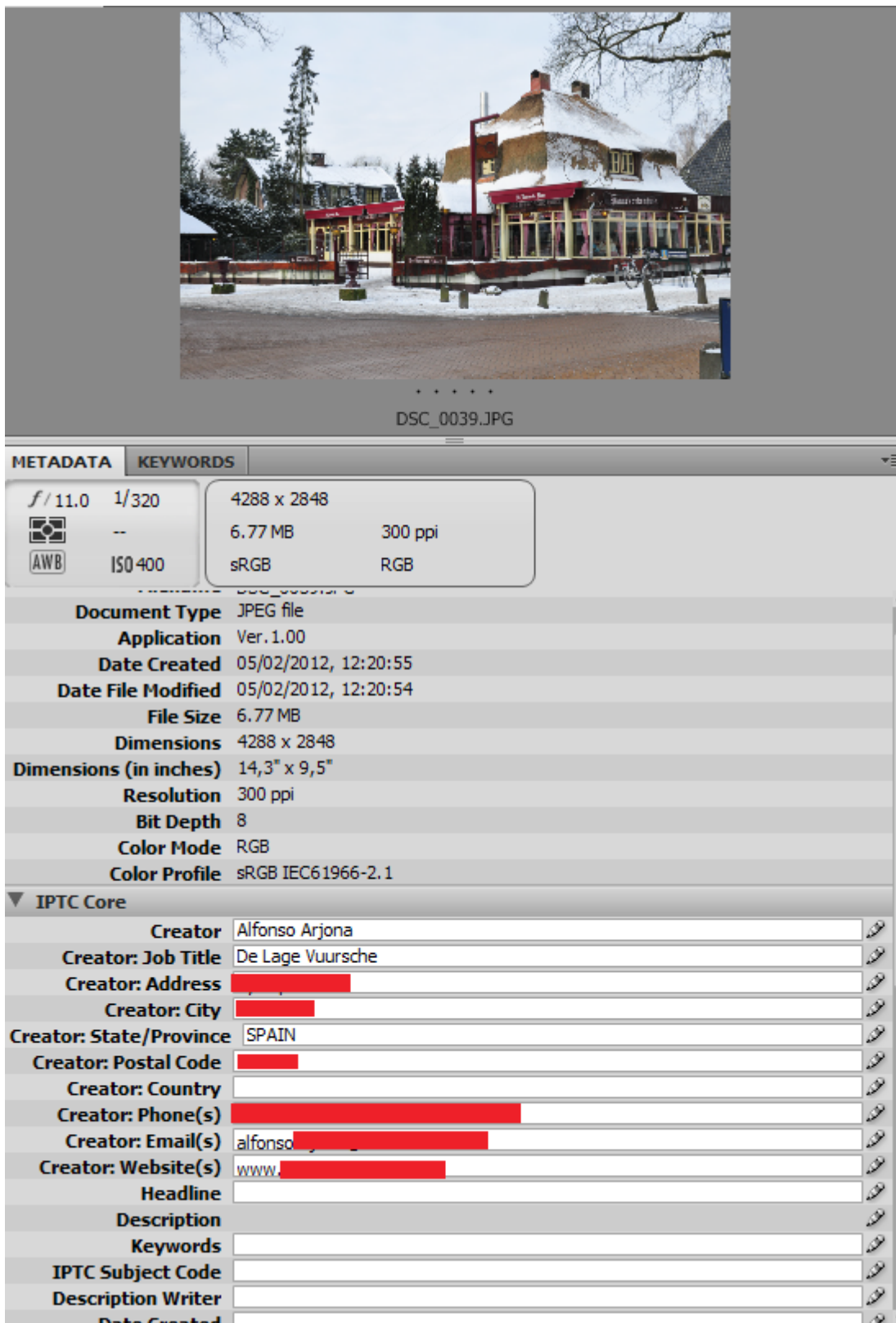
- 15.6 Using LibreOffice Writer, select File | Properties and see what you can learn. Open a web page and "view source" in the browser. The metadata is usually at the top of the file. Within the page, you may find comments, variable names, or indications of client-side scripting. Within Microsoft Word, select File | Properties or File | Prepare | Properties. Select a Microsoft Office document and open it in a text editor. You will find metadata in the header and in the end of the file. What can you learn?

Within an organization, information is shared with people who have direct control over that information or are influenced in any way by that document (not really, I just wanted to see if you were paying attention). For example, the head big wig with a big belly and thinning hair wants to change the font the entire organization uses because his cat likes that font. First, the legal department is contacted to see if there would be any legal ramifications within the law and font ownership rights (trademarks). Next, the Human Resources division is called in to see if there would be any backlash from the local unions. As you can see, many folks are part of a single whim and have to deal with all the footwork. Along the way, new information is added, edited, taken away, remitted, highlighted, trashed, reworked, and such.

The real jewels for you are the names of people who had their hands on certain information and when. Departments have to talk to each other otherwise they will just be another U.S. government agency. A comment made in the footnotes by the paralegal foot soldier about possible conflicts with a new fiber optic upgrade in one document may help you add another piece to the network puzzle. Another blast on a blog by some company IT woman complaining about the long hours she is working to help get the IDS working correctly, can also provide you with critical intelligence on your target.

The origins, contents, and destination of specific documents can provide a framework to profile organizational personnel who don't get any credit for their hard work, yet are privy to valuable inside information.

If you take a picture with your digital photo camera, the file will contain EXIF information. This data includes ISO values, aperture, focus and information about the shot, but also GPS coordinates or even your name, e-mail or home phone number.



That's what we call **metadata**, and it can be found on photographs, video, music files, web pages or similar, but also in PDF, DOC, ODT, TXT, XLS, PPT, RTF, MDB files and so on. This



gives an attacker extremely valuable information about you, your preferences, and your contact information and can be used to gain access to the system. Does this mean metadata is a weak point for security? Absolutely not.

Imagine yourself 10 years in future, a famous film director. Sounds great. Private parties, beautiful people everywhere, tons of money and filming a new blockbuster: Revenge of the Space Octopus XVII. But...wait! This is not the right screenplay! The whole scene is wrong! (It's time to run in circles waving your arms and yelling) Fortunately, the producer is in a "business trip" to Bora-Bora, so you just need to film the scene again with the right version of the screenplay before he comes back.

But when you go to the office to print a copy, you forgot the version number. You have too many revisions of the screenplay! (Everyone knows a space octopus movie requires very philosophical dialogue.)

No problem. You still remember that this dialog has been written the 4th of July, at New Jersey Island, last summer (I know you did it). So, you decide to perform a search using those keywords and... Bingo! Document found! If you did not include the metadata, you could not find the document.

So, protecting your privacy is not a question about to include metadata or not on documents. Is just being careful about where the document will be stored, and who is allowed to use it. You may wish to search for and use tools to strip the metadata from your files before you post them or send them to someone.

Exercises

- 15.7 Search online for .jpg, .doc or .xls documents related to a site. Then have a look at the properties of the file. Can you find sensitive information? (Tip: use "site:whatever.com filetype:doc" to refine the search. This will show you .doc files at whatever.com.)
- 15.8 Open a photograph you have taken, a document file containing your homework or a music file, and have a look to the properties. Can you find information about you or your computer?
- 15.9 Browse the web and find pictures people have posted, e.g. to Facebook or a personal blog. What metadata can you find in the pictures? You can use File | Properties and look at the Summary tab in Windows. You can open the file in a text editor and search through it for recognizable strings of text. Within Linux/UNIX there is a "strings" utility that will search out and display all the ASCII strings in a file.

P2P Networks

This is a point to discuss seriously.

First of all, there is nothing wrong with P2P or direct download servers. It's a very helpful technology used worldwide. You've probably heard complaints against P2P, but you have to remember it's just a protocol. The content that is being distributed is a different story, but you don't ban trucks become some carry drugs.

Ok. Let's go.

Sharing Documents

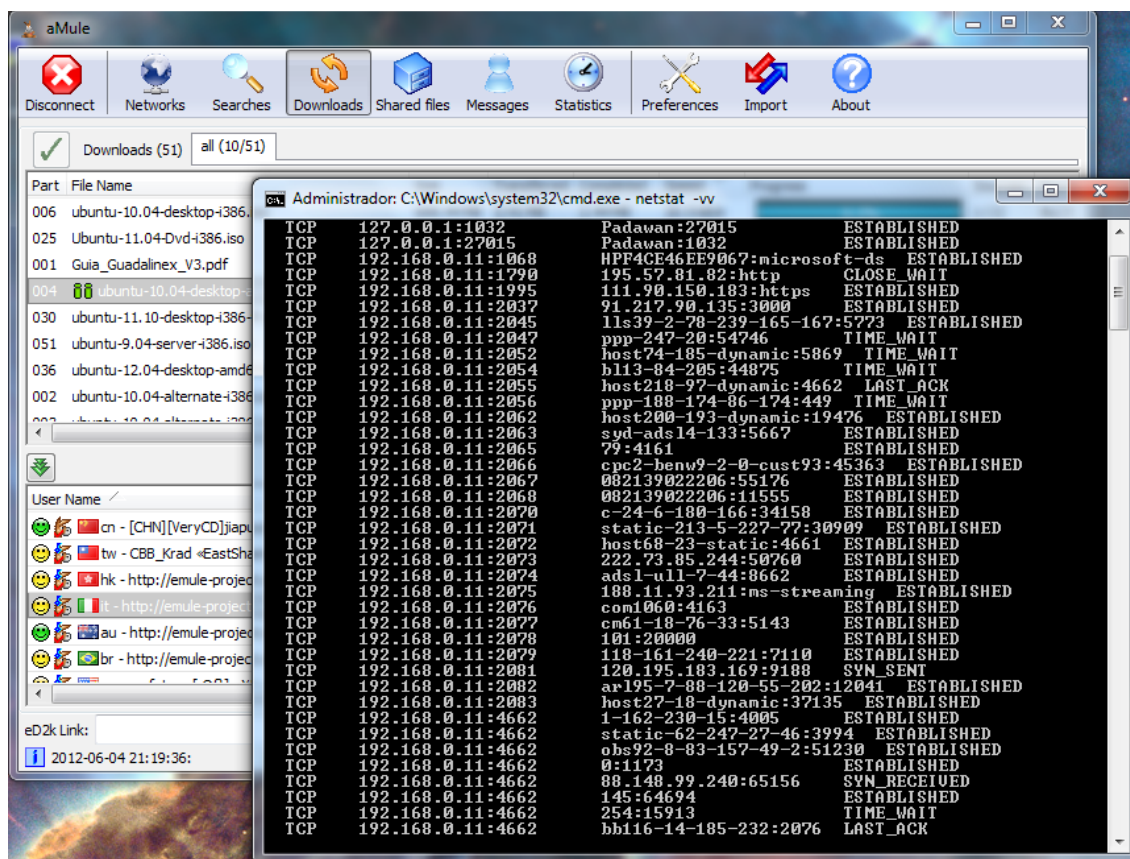
Did you know several companies are using P2P networks for distributing contents? Some **MMORPG** games use P2P to improve the distribution of updates. Others use it to share ciphered files containing templates, documents or databases . Even some governments

use P2P to distribute operating systems, applications, and push software updates. Don't believe it? Search for **Guadalinux**. It's a free operating system created in Andalusia (Spain) used in governments, schools, universities, public libraries and more, and used worldwide by Spanish speakers.

If a hacker knows that a company shares documents using P2P, the first target is to obtain the IP addresses that are downloading those files. Why? Because some of those IP addresses will be owned by the home office, others by people working in a different office (or even at home) and the rest of them other people around the world who think the file contains a pirated movie, song or whatever.

How can those IP addresses be found? Quite simply. First of all, use a P2P client to share the latest Guadalinux ISO file (it's open/free software, so everyone is allowed to do that).

After a while, open a terminal and type "netstat -vv." This will list all connections to your computer.



The first column shows the protocol, the second one is your private IP and the third one shows the IPs used by the people downloading the file from you.

After having a look to the client used to share the file, you can see the name set on the remote IP. Most times this will be the default user name for the application, but in a corporate environment it's usually a name like jimkirk@unitedfederationofplanets.org, an email address, which means that jimkirk may be the username to log into the system.

Simple, isn't it?



Fake Files

Another way hackers get into systems is to use a Trojan. This is something quite easy to do in P2P networks, sharing a **fake file**. A fake file is a file with the same name as the real one, but it contains malware that will be installed on the victim's computer when they decompress or execute the file. This allows a hacker to control the computer, log everything the user does, and save this information for his purposes. This fake file should be the same size as the real file. If you label a fake file "Office 2010," and it is only 10k in size, there is a fairly good chance that it will be recognized as a fake. Looking further down the road, check sums and MD5 hashing could be compromised too so the fake file looks like the real thing.

Imagine this situation: your high school shares lessons using P2P networks. Most of them will have exciting names as "Chemistry, lesson 4," "Geology of the Rocky Mountains IV" or "Mathematical study to calculate the Fourier transforms." Also, there is a file named "Practice test for Chemistry," and a self-extracting file named "Chemistry: final test with answers" Wow! Looks awesome! You can take the exam and have an A+! It's cheating, but no one will realize, unless your teacher planted those files. So, you decide to download the "final test with answers." Once the download is finished, you execute this file and find that the content is the same as the other practice test file. The difference is that now your computer has a Trojan, and your privacy is gone.

You can use file properties, metadata, digital file signatures, or hashes or checksums to compare downloaded files to the "same" file from another source and decide if you have a real, or a poisoned (Trojaned) file.

Former Employees

There is an amazing fact about doing research: Did you know most of the information disclosed about a company on the Internet has been exposed by ex-employees? As a wise Master said "Fear is the path to the Dark Side. Fear leads to anger, anger leads to hate; hate leads to suffering. I sense much fear in you."

Being fired is not nice, but revenge can get people in serious trouble. Despite knowing that doing so can get them in legal trouble, some people will release confidential information about a company via P2P. Usually, those files will contain usernames, passwords, phone numbers, financial information, compromising documents and photographs...all the kind of information a hacker hungers for.

Password Protected Documents

Don't think documents owned by a company are shared without protection. They may use a strong password to keep people from reading the content. Even using a password cracker, if the password is a long string it will take years (or centuries) to find it. But sometimes, if the password is not too complex or there is a bug in the encryption tool, the file can be decrypted faster, maybe in hours or days.

That's why a hacker will also look for old files. Even if the file is several months (or years!) old, it may contain useful information regarding the target.

Exercises

- 15.10 Install a P2P client, and perform a search about yourself, your high school or your city. Can you find shared files that may contain information?



- 15.11 Share a non-copyrighted file (a Linux distribution is a good choice). Examine the connections made to your computer.

Search Engine Hacking

In social engineering, information is everything. Knowledge is power. The better you know your target (person or system) the better you can manipulate it. Search engine hacking is about collecting intelligence openly available on the Internet and other sources. Intelligence services call this discipline open source intelligence (**OSINT**).

New hackers and security experts have different approaches, tools, and skills to perform their duties. One of the tools used by both is search engine hacking. Of course, search engine hacking can also be used for espionage or targeted attack preparations. However, this requires a higher degree of technical knowledge.

Within the hacker and information security scene the term **Google hacking** is often used to describe the methods and procedures to obtain OSINT from the internet. Of course, Yahoo! and Bing can be used or misused in the same way as Google, so the generic term search engine hacking is more correct.

With search engines, you will find almost everything, or at least everything their robots (search spiders) collected. Quite often, the robots crawling the World Wide Web find and index information not meant for public viewing. From a business and user's perspective, this is a hidden security risk. Millions of indexed pages offer a variety of vulnerabilities. Anyone who has ever "Googled" his own name wondered what he could learn about himself. If you are looking for clues or references to your own name on the Internet, this is referred to as **ego surfing**.

What kind of information can you find by searching?

Configuration Files

Configuration files reveal sensitive information like user names and passwords. Configuration files can often be found with file extensions like .ini, .conf, .config, .cfg or .dat.

Log Files

Log files can reveal sensitive information such as IP addresses and user tracking. Log files can often be found with file extensions like .log.

Office Documents

Documents that are unintentionally posted to public areas can contain sensitive information, including metadata. Common file extensions include .pdf, .doc, .txt, .xls, .odf, .odt, .ods, etc. Document names like private, password, backup, or admin can indicate sensitive documents.

Portals, Databases, and Default Hardware Passwords

Login portals, especially default portals supplied by the software vendor, act as magnets for attackers. Often these portals and standard hardware complements still have the default (well-known) installation passwords on them.

- The words "login," "welcome," and "copyright" are excellent ways of locating login portals.
- Support files exist for both server and client software. These files can reveal information about the configuration or usage of an application.



- Error messages have varied content that can be used to profile a target.

Database dumps are perhaps the most revealing of all finds because they include full or partial contents of a database. These dumps can be located by searching for strings like "# Dumping data for table."

Some administrators configure their websites so badly that no exploit is required to gain access to the system. Search engines index the web very aggressively and discover virtually any file on a website – unless it is in a password-protected area or otherwise secured. The wrongfully or accidentally indexed information can include: password files, credit reports, health data, etc. In cases where the files are not adequately protected against indexing, the search engine has basically executed the attack on your behalf. The data can be used as raw material for targeted phishing scams and other malicious activity. The barriers to search engine hacking are low since you need next to no technical skills.

How Search Engines Function

By default search engines are built for speed so they skip the so called **stop words** in the search query. Stop words are small, short words like "the" and "to" and "and." Google will ignore these words in a search. Also special characters like @ # \$% ^ & * (\) = +] are usually ignored unless these are well known terms that have significance, such as "C++" or "C#". For example, if you type in the string *the who*, "the" will be ignored. However, you may want to search for the group The Who. In order to do this, you need to use quotation marks, i.e. search for the string "The Who", quotation marks included. You can also put a plus sign (+) in front of the search term, like "+The Who."

Basic Operators

With the plus sign you can join several terms so they are linked together. This is equivalent to the standard Internet search, i.e. *blog + comments*. You don't have to write out the plus signs; the search engine automatically inserts them into the spaces.

(-) The minus sign is used for exclusion i.e. "Hacker High School -boring" only returns results with Hacker High School, but without results having boring included.

(|) The pipe symbol is used to separate terms (either one or the other), i.e. "Hacker High School + fun | exciting." This search query is looking either for Hacker High School combined with "fun" or "exciting."

(" ") Double quotes search for the exact phrase, i.e. "John Ken-nedy" is searching for John Kennedy but not for John Fitzgerald Kennedy.

(~) The tilde symbol looks for similar terms or synonyms i.e. "~ Hacking" is looking for terms similar to "hacking."

(*) The asterisk is a wildcard search. With the wildcard you can replace words that are not known, i.e. "Pizza, without *"

(..) Two dots (periods) do a "from - to" search. This can be used if you are looking for a laptop between 400\$ and 600\$ you would search "Laptop USD 400..600".

Advanced Operators

filetype: With filetype you can search for specific file types i.e: "filetype: pdf" This will search for pdf files for instance.



intitle: Lets you search the <title> tag. For instance "intitle: index" will find you documents with the word "index" in them.

intext: With intext: you can find words on a website.

For instance "intext: Hacker" will find you documents with the word hacker in them.

inurl: With inurl: will find words within a URL.

For instance "inurl: etc" or "inurl: bin" will find you "etc" and "bin" folders.

site: With the site: parameter you can search on specific domains.

For instance "site:com" or "site:gov" will find you results on either .com sites or .gov sites.

cache: With cache: the site is displayed, as it is inside Google's cache. For instance "cache: www.blick.ch"

related: With related: Pages are displayed which are similar. Example: "related: pizza"

safesearch: With safesearch: will prevent adult web pages or adult content from being displayed. Example: "safesearch: Pamela Anderson" will show you only safe results when searching for Pamela Anderson (there may not be many).

The Dark Side

Where is the threat? If you look at the fundamentals of successful searching, it's a small step to use this knowledge in an unethical or even unlawful way. Searching for vulnerabilities with the help of a search engine does not necessarily lead to a criminal offense. The search engine is not the culprit or hacker; it's only the discoverer of information. A hasty criminalization is not appropriate. However, the offense is based on a search engine search with the help of technical on-hand witness factory initiated a breakthrough experiment. The offense is based on what you do with the results.

Useful Tools

What else is possible with Google?

Weather: Provides disclosures about the local weather Example: "Weather Washington"

Calculator:

Ex: "(5 * 9 + 3) ^ 3" or "20% of 500"

Conversion of units: Ex: "25 Miles to km" or "0xFF - 0b11010101 to decimal"

Currency Conversion: Ex: "50 € in USD"

Definitions:

Example: "define: nomophobie"

Make Web Searches Easy: Customize Them

The following website offers you the opportunity to compose your own search queries in a simple manner.

Reputelligence™ search works on the principle: Go out there and find me the stuff that I'm interested in quickly, and do it in a way that is precise and repeatable. Contemplating the evolution of our use of the Internet over the past 20-odd years makes one aware of the different stages and expanding scale of information accessibility. In the early days, the Internet was something "over there" that didn't really concern us. Then everything exploded. "Dot coms" happened, and the barrier to getting online began to come down



to the point where it wasn't just Big Corporations that could have an online presence. Now, we're about to move into a new phase of the expansion of the Internet: the search for relevant content, and its organization and presentation to the user.

There's too much content available and too little time to consume it all. Most of us just want a way to sort the wheat from the chaff, the gold from the dirt. Reputelligence has been purpose-built to tailor the user's search experience. It is a way of constructing comprehensive search queries without the need for deep search string knowledge. Reputelligence helps you to craft comprehensive search queries, creating a shortcode URL and a QR code of the search and executing it for you.

The shortcode URL and QR code is a very useful feature, especially when people ask, "What have you been searching for?" or "How did you find that?" Simply send them the shortcode URL or the QR code, and they can have the same search experience as you had. This approach makes search results more relevant, more specific, and more targeted, and the user receives results that are more accurate in less time. Check it out and enjoy!

Exercises

Who wants to try out a few searches? Web cameras can be found with the following strings:

inurl: ViewerFrame fashion =?

inurl: ViewerFrame Mode = Refresh

inurl: axis-cgi/jpg

inurl: axis-cgi/mjpg

inurl: view / indexFrame.shtml

inurl: view / view.shtml

15.12 Use this information to locate web cams in London, New York, and Taegu, South Korea. See if you can find anything interesting in these cities.

15.13 Locate other people that have your last name and live in your area. You can also try and locate people that have your last name in London, New York, and Madrid (not Korea, nobody has your last name in that country). See if they have a web cam, too.

15.14 Don't be a stalker, but see what information you can find on the person who you are most interested in. You might want to look up some cute person in school, your favorite musician, that one librarian that keeps looking at you like you're going to steal a book, or your favorite sports star/writer.

Today's teens are in a world with major communication and productivity channels open to them and they don't have the knowledge to defend themselves against the fraud, identity theft, privacy leaks and other attacks made against them just for using the Internet. This is the reason for Hacker Highschool.

The Hacker Highschool project is the development of security and privacy awareness learning materials for junior high and high school students.

Hacker Highschool is a set of lessons and a practical means of making hackers. Beyond just providing cybersecurity awareness and critical Internet skills, we need to teach the young people of today how to be resourceful, creative, and logical, traits synonymous with hackers. The program contains free security and privacy awareness teaching materials and back-end support for teachers of accredited junior high, high schools, and home schooling. There are multiple workbooks available in multiple languages. These are lessons that challenge teens to be as resourceful as hackers, including safe Internet use, web privacy, researching on the internet, avoiding viruses and Trojans, legalities and ethics, and more.

The HHS program is developed by ISECOM, a non-profit, open-source research group focused on security awareness and professional security development and accreditation.