

# Offensive and Defensive Techniques for Determining Web User Identity

- Zachary Zebrowski
- [zakz@zakz.info](mailto:zakz@zakz.info)
- Approved for Public Release: 12-3046.
- Distribution Unlimited

# All materials are licensed under a Creative Commons "Share Alike" license

- See <https://creativecommons.org/licenses/by-sa/4.0/>

# Welcome

- I'm Zak Zebrowski.
- I'm a Forensic Database Engineer / "Data Miner" / and Perl guy.

# Outline

- Characteristics of internet connections
- Ways of determining web user identity
- Preventing a web user identity from being determinable from an end user's perspective
- Finish

# Ethics

- Reminder: Use for good not evil.
- I am not a lawyer.
- What you do is at your own risk.

# Why Bother

- To determine who visits your website.
- Validate you are connecting from a known location versus an unexpected country.
- To hide your identity on the internet
- Note: Everything in this presentation is old. New technologies and techniques evolve over time which may be more (or less) intrusive.

# Characteristics of Internet Connections

- What devices connect to the internet?
- When you connect to the ineternet.
- How you connect to the internet.
- Why bother : You can't escape these characteristics regardless of what you do.

# What devices connect to the internet?

- Easier to say what devices don't connect at this point. Largely, things that don't have power. Most everything else, like your car, ebooks, refrigerator, dishwasher, etc, especially newer models with higher end features, can and do. (I just got notified my dishwasher got done washing dishes.)

# When do you connect to the internet?

- Most people go to sleep.
- You can get an approximate location of someone, by noting times when people login or not.
- See [https://census2012.sourceforge.net/images/geovideo\\_lowres.gif](https://census2012.sourceforge.net/images/geovideo_lowres.gif) . This is an image of the (relative) number of devices that were turned on over the course of the day in 2012. Note that the number of devices go down (denoted by the light blue color) over nighttime and the number goes up (orange red) over the course of a day.

# How you connect?

- If you think outside of the modern economy, not everyone has a cell phone... Or if they do, it may be slower performance or faster performance depending upon where someone lives (city versus country... and different countries as well with better economies.)
- Also, different regional differences (such as hurricanes, where infrastructure is not reliable) could redesign the average connection to the internet.

# Other identifying techniques

- IP Geolocation
- What your browser exposes when you connect to the internet

# IP Geolocation

- IP Geolocation is the name for the technology to go from an IP Address to a physical location. Companies, open source projects, provide this information in various formats to download and use within your applications (internal or external). Further companies can provide API's as well to various websites through techniques.
- Note that companies often provide additional attributes as well as IP geolocation, such as domain name, isp, org, and other calculated attributes. These are a by-product of the analysis they perform to refine calculations.
- As location information is valuable, other projects attempt to refine location as well, through bluetooth or wifi devices, cell phone coverage, or other attributes and then allows the browser to perform an approximate location lookup based upon the devices found and their known locations. See <https://beacondb.net> as an example data source. Applications (such as firefox, or geoclue), can make requests to this api for an approximate location response.

# What are unique features about IP Geolocation Databases

- An IP Geolocation database is often offline. As such, I can tell from logs, approximately where you are located, without performing online queries.
- Though IP Geolocation, combined with other resources, I can generally identify your ISP provider (verizon or similar) and then target you based upon that information. I could also potentially identify fraud, if your profile does not match where you are expected to login from.

# How does one create a IP Geolocation database?

- (Caveat, only open source IP Geolocation techniques are discussed.)
- Ask the user. Hostip.info was (past tense) a source that simply asked the user, for a given IP Address, what country they were located in.
- Query registerys. Registrys assign IP Addresses to countries, and provide files of IP addresses to ISP. Generally, these ISPs are known to operate within a given country. Software77.net (a now defunct website) did this, as well as ipasn.com , which provides ip to country information.
- RFC 8805 (and supplement 9632) specifies an open way for ISPs to designate locations for ip address ranges. The supplement specifiys how to identify the locations of the 8805 files for a given ISP.

# Differences between open and commercial sources

- Open sources have known algorithms and reues.
- Commercial sources: Claims of higher accuracy, and dedicated staff to further look into the data.

# IP Geolocation Caveats

- You can get only to the most visible IP Address.
- Generally IP Geolocation files are delayed, and then need to be ingested into your environment. This may take time.
- There is generally a 1-2% drift per month for location information over time.

# IPv6

- IPv6 addresses are used in various countries, and its use is being increased over time.
- At the ISP level, IPv6 is used a lot to help with internal networking.
- Some consumer level IPv6 connections exist. It depends upon the provider.
- IPv6 can be translated to IPv4 communication, through transparent proxies, sometimes at the OS level or at the ISP level.

# Lab 1

- Browse to <https://location.zakz.info> and click on the Lab 1 link.
- Expand the summary tab Server Information to see all of the server information that is collected upon each request. Some of the information is the server environment proper, such as what web server is running, etc. See if you can find other identifying information.
- Note that without any additional permissions, the server can (reasonably) guess what country you are located in. The server can also identify (reasonably) the network connection (company) that you are using, as well as to if you are or are not using tor.

# What your browser exposes

- By default, your browser exposes your IP address or your proxied IP address to the remote server. In lab 1, you can open the server environment information tab to see the REMOTE\_ADDR field containing that IP Address.
- It also exposes your User Agent (which is what web browser you use). Because of privacy concerns, the user agent string is becoming more static over time, to reduce tracking opportunities via the string.
- It also exposes cookies that the web site you are visiting has sent in the past, (which may contain session information or other information to get you quickly back into your account). Cookies are a key value pair that is returned by the browser to the server with every request to that domain. Cookies can be used across domains to track you over a longer period of time.

# What is exposed by scripting

- Javascript (and other scripting languages) can provide insights into your behavior.
- Look at [whatwebcando.today](http://whatwebcando.today) website. It shows you many APIs that your web browser can use to interact with your environment.
- Some of the interesting APIs: Audio & video capture, Geolocation, Device motion, Network Speed, Online State, Battery Status, Touch Gestures, clipboard, etc.

# Lab 2

- Browse to whatwebcando.today/geolocation.html website.
- Click on the "ask for location" button under Live Demo.
- Click allow location when prompted. (By default, websites must ask).
- Click on the coordinates that were fetched by the api.
- See how physically close the location that is returned to your own location.

## Lab 3

- Browse to whatwebcando.today/battery-status.html
- Note that you do not have to click anywhere additionally to see live demo information about you.
- See the live demo, particularly on a cell phone.
- 

**At one point, Uber was accused of increasing ride prices for individuals based upon this information, however, that was incorrect. See a news report here : <https://preview.tinyurl.com/bavvu2r6>, that talks about this in detail.**

## Lab 4

- Browse to <https://panopticlick.eff.org> and run a test.
- The test will take a minute or so to run.
- When results are available, scroll down to Fingerprint Metrics.
- 

**Note the first line is System Fonts. As you read through this, ask yourself if your company has a company font installed on its browser. If it does, it could potentially identify your visit from your company, regardless of other obfuscation techniques your company may employ to avoid having the public know that a person from your company is interested in it.**

## Lab 5

- Browse to <https://preview.tinyurl.com/3yqbptk>. This will redirect you to a tech dirt article, written in 2010, about a technique that used to work. A script was installed on youporn website, which allowed the website to determine what other websites your browser visited. It used the default color of a href link that is visited or unvisited, to

determine if you went to a competitor's website.

- While the technique no longer works, it's worth noting that companies are interested in what sites you have visited before. It would be likely that there are additional techniques people are working on to create this history for you.

# Lab 6

- Browse to <https://dugwood.com/clickheat/>
- This is a javascript application that allows a web site owner, to see where visitors are clicking on their website, in a heat map fashion.
- This is another technique that one could use to potentially identify you.

# Lab 7

- Browse to <https://github.com/samyk/evercookie/>
- Evercookie is a javascript library that provides persistent cookies in a browser, using multiple techniques, that may be hard for the user to control. Clearing cookies is not enough to get rid of evercookie.
- The current version is sort of frozen in place, but some of the techniques still work. Additional techniques are being developed. Look at the issues and pull requests to see ideas / issues with current evercookie library.

# External connections

- It's possible to embed multiple connection types (http, ftp, smb, etc) through a web browser. By using obscure protocols, they may "break out" of your protected web browser session.
- See <https://preview.tinyurl.com/brcg6ca> - a defcon 17 presentation called Attacking Tor by Gregory Fleischer that has additional details on how to escape from a web browser (including tor). While dated (8 years old at the time of writing), it's a little dated at this point, however, the principles do not change.

