

Offensive & Defensive & Forensic Techniques for Determining Web User Identity

Part 1

Zachary Zebrowski
zak@freeshell.org

Approved for Public Release: 12-3046.
Distribution Unlimited

All materials is licensed under a Creative Commons “Share Alike” license.

- <http://creativecommons.org/licenses/by-sa/3.0/>

You are free:



to Share — to copy, distribute and transmit the work



to Remix — to adapt the work

Under the following conditions:



Attribution — You must attribute the work in the manner specified by the author or licensor (but not in any way that suggests that they endorse you or your use of the work).



Share Alike — If you alter, transform, or build upon this work, you may distribute the resulting work only under the same, similar or a compatible license.

Welcome!

Let's introduce one another...

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



Class Outline

- Introduction
- Characteristics of connecting to the internet
- Internet Networking Background
- Offensive Ways to determine a web user identity
- Defensive ways to prevent determining a web user identity from the end user's perspective
- Additional Tasks from those above
- Forensic Database Analysis
- Forensic Web Log Analysis
- Finish

Introduction

- Why bother?
 - To determine who visits your website on the internet
 - Your bank wants to know if I'm in the US, or in the UK
 - To hide who you are on the internet.
 - To determine forensically who visited your website after some event happened.
 - Cool! I want to be an 31337 H4x0r
 - Not exactly. **Everything in this presentation is old.** New techniques evolve over time.

Ethics

- Reminder: Use what you learn here for Good not Evil.
- I am **not** a lawyer.
- What you do is at your own risk.

Characteristics of Internet Connections

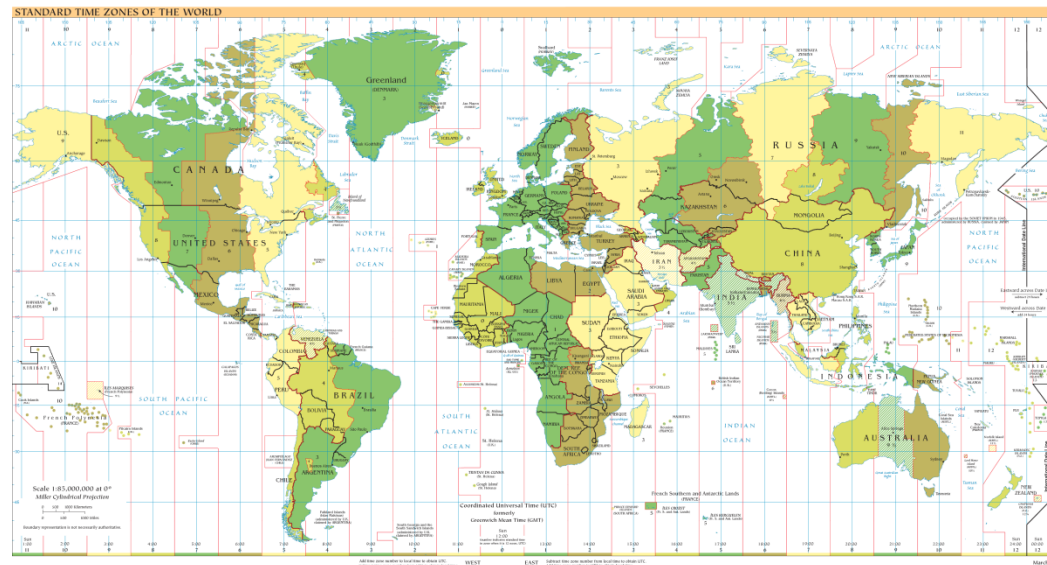
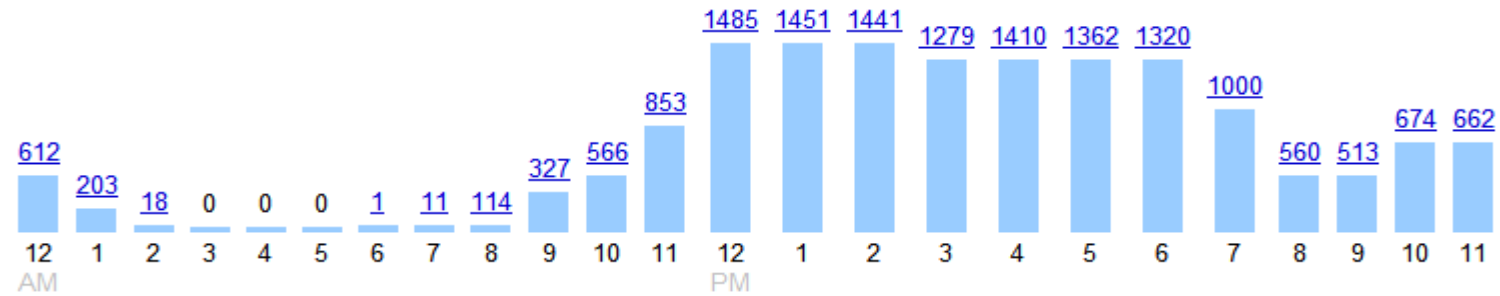
- Questions to ask:
 - What devices connect to the internet?
 - When do you access the internet?
 - How are you connecting?
- Why bother?
 - *You can't escape these characteristics, regardless of what you try to do...*

What devices connect to the internet?

- Almost easier to say what device doesn't connect.
- Here's a list of things known to connect to the internet:
 - Computer
 - Cell Phone
 - SIP Phone (VOIP)
 - Apple TV / Google TV
 - Your thermostat (Multiple companies)
 - Your alarm clock (Chumby)
 - Your car (Chevy Volt)
 - Your door lock (Schalage)
 - Your picture frame
 - Your camera (EYEFI sd card)
 - Your watch (pebble ; impulse ; etc)
- && Each of these devices can be identified through various means
 - `nmap -O -V 192.168.1.1`

When do you connect?

Hourly search activity



Source: google.com analytics (for my personals search history) && <http://en.wikipedia.org/wiki/File:Timezones2010.png> (licensed public domain).

How are you connecting?

- Dial Up
- Cellular Phone Device
- DSL / Cable
- FIOS

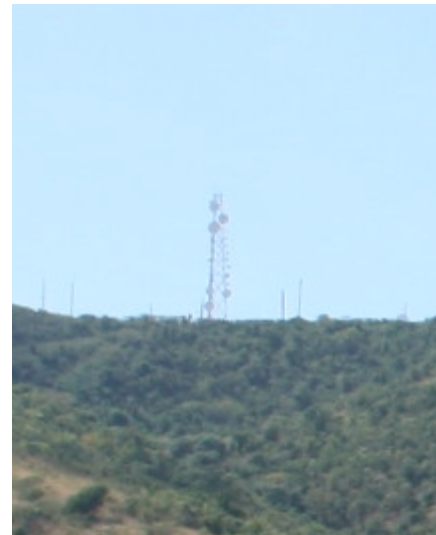
- Other ways:
 - Satellite connections
 - Radio connections

Satellite Example

- Designed for remote areas where other options not available
 - See <http://www.wildbluesales.com/> as an example
 - As low as \$39.95 / month ; 1mbs download ; other plans available
 - Valid as of 6/28/12

Radio Connections

- Found in the US Virgin Islands / PR
- Mountainous terrain; good weather (minus hurricanes); fixed coverage area; somewhat poor utility infrastructure.
- <http://www.ackley.vi/>
 - Photo source personal photos



Characteristics of Internet Connections

Summary

- What devices connect to the internet?
 - Almost anything, but they're detectable
- When do you access the internet?
 - Can generally detect what time zone you're in.
- How are you connecting?
 - Can be summarized if speed is detected.
- Why bother?
 - ***You can't escape these characteristics, regardless of what you try to do...***

Basic Internet Technology Background

- What is an IP Address?
- How are IP Addresses Assigned?
- What is a NATed / Private IP Address?
- What is a Port?

What is an IP Address?

- An IP Address is a unique identifier that allows you to connect to the internet.
 - Conceptually, it's similar to your street address for your house

What is a NATed / Private IP Address?

- There are a limited number of IP Addresses available. A ISP may assign you a particular IP Address, but not enough for all of your personal devices. “NATing” allows you to have many personal devices, while using only one public IP Address on the internet.
 - Conceptually, it’s similar to an apartment number in an apartment complex.
- A private IP Address is simply a non-routable IP Address on the internet, which the home router gives to local machines, and via NAT routing connects to the internet.

How are IP Addresses Assigned?

- IP Address ranges (a set of IP Addresses) are assigned to an ISP (Internet Service Provider) by a registrar. The specific registrar is dependent upon what domain you are purchasing for, though generally this is treated based upon the region in where you live.
- A particular ISP can then assign you a particular IP Address for your router.
 - But the private IP Address space is only assigned by your router.

What is a port?

- A port is simply a number, which allows two computer programs on different computers to communicate with one another.
- Generally, certain services run on specific port numbers, however, a programmer can arbitrarily set what ports to connect to if there is a good reason too.
 - Port 22 for ssh traffic
 - Port 80 for http traffic

Questions?

- You just survived the “boring” part.
- 5 Minute Coffee / Bathroom break.
- When we come back, we will start with offensive techniques to determine who a user is.

Identifying Techniques

- These are server techniques to determine who you are.
- IP Geolocation
- What your browser expose when you browse to a web page.

IP Geolocation

- IP Geolocation is the name for the technology to go from an IP Address to a physical location. Companies, and open source utilities, provide this IP Geolocation information in various formats to download off of the internet for use within your internal applications.

Note: Sources generally provide other attributes *as well* as just ip location information, such as domain name, isp, org, or similar. This is a by-product of the analysis they perform.

IP Geolocation – what does that mean?

- For a given IP Addresses, I can tell you approximately where they are located, *without* using a network connection.
- Also, I can generally tell your internet provider
 - Verizon.com
 - Yourbuissness.com
 - Etc
- I can use this to:
 - Send you targeted internet ads
 - Verify you have permission to view media content
 - Verify you are in a country the same as your credit card

How does one create an IP Geolocation database?

- ▶ *Caveat: These only discuss how open source IP Geolocation techniques work, not commercial sources.*
- Ask the user
 - Hostip.info
 - Host name: **pool-70-108-49-201.res.east.verizon.net.**
IP address: **70.108.49.201**
Location: **UNITED STATES** ([change](#))
 - <http://www.hostip.info> (12/18/10, 1/7/12)

How does one create an IP Geolocation database (2)?

- Query registries
 - We [Software77.net] use the registry assignments provided by the registrars. However discrepancies creep in especially in cases of large multinational companies who have their base of operation on one country and satellite offices in other countries. Typically what happens is that a company based in say, the United States, also has a branch in Africa or Asia.
 - <http://software77.net/faq.html> (1/24/08)

Open Sources

- Maxmind
 - Country & City files (less accurate than paid version)
 - <http://www.maxmind.com>
- Hostip.info
 - Community volunteers location, plus various automated features
 - <http://hostip.info>
- Software77.net
 - Scrapes whois notifications of new domains
 - Provides a “birth date” for a domain.
 - <http://software77.net/cgi-bin/ip-country/geo-ip.pl>

Open Sources VS Commercial Sources

- Open Source
 - Less data fields
 - Known algorithms
 - Unique attributes
 - Hostip.info is user contributed – could be interesting?
 - Software77.net has *daily* location updates.
 - Free
 - Convenient if you need to do this analysis infrequently
- Commercial Data Sources
 - Claimed Higher accuracy
 - Dedicated staff
 - Costs Money

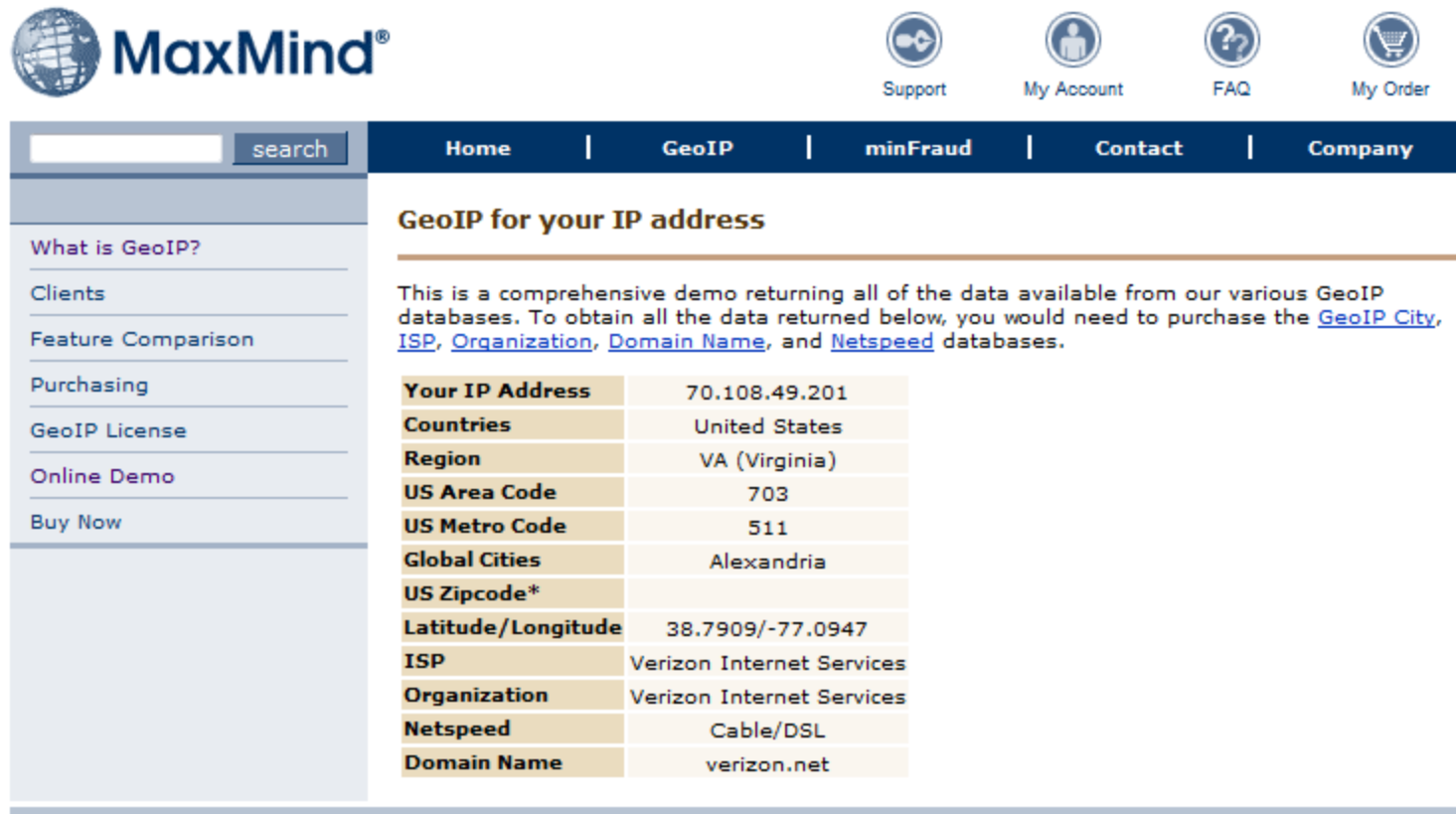
IP Geolocation Caveats

- Can only get to the most visible IP Address.
 - No indication that a proxy is being used
 - (unless a well known proxy)
- Time frame
 - Generally the data is delayed at least one month, plus time required to get the data into a usable format.
 - IP Addresses can be re-allocated from ISP to ISP in a short period of time.
 - Newly registered host names may not show up in a short amount of time
 - About 1-2% change per month (<http://www.maxmind.com/app/faq>)
- Domain name resolution is limited
 - Multiple domain names can point to a single IP Address
- “NATed” IP Addresses
 - Where a user at home uses a broadband modem to connect externally to the internet. Behind the modem there may be many machines, but there is only one IP Address facing the internet.
 - Alternatively, could be a rural ISP provider.

IPV6

- Starting to be integrated.
 - Currently “a small percentage” in any country hosts IPv6 servers.
 - [http://www.ripe.net/ripe/meetings/ripe-57/presentations/Colitti-Global IPv6 statistics - Measuring the current state of IPv6 for ordinary users .7gzD.pdf](http://www.ripe.net/ripe/meetings/ripe-57/presentations/Colitti-Global%20IPv6%20statistics%20-%20Measuring%20the%20current%20state%20of%20IPv6%20for%20ordinary%20users%20.7gzD.pdf) or <http://preview.tinyurl.com/b36qzo>
 - <http://www.maxmind.com/app/ipv6>

Example of information returned for my home connection in **Arlington, VA**



The screenshot shows the MaxMind website interface. At the top left is the MaxMind logo. To the right are navigation icons for Support, My Account, FAQ, and My Order. Below these is a dark blue navigation bar with links for Home, GeoIP, minFraud, Contact, and Company. A search bar is located on the left side of the navigation bar. The main content area is titled "GeoIP for your IP address" and contains a paragraph of introductory text followed by a table of IP-related data.

GeoIP for your IP address

This is a comprehensive demo returning all of the data available from our various GeoIP databases. To obtain all the data returned below, you would need to purchase the [GeoIP City](#), [ISP](#), [Organization](#), [Domain Name](#), and [Netspeed](#) databases.

| | |
|---------------------------|---------------------------|
| Your IP Address | 70.108.49.201 |
| Countries | United States |
| Region | VA (Virginia) |
| US Area Code | 703 |
| US Metro Code | 511 |
| Global Cities | Alexandria |
| US Zipcode* | |
| Latitude/Longitude | 38.7909/-77.0947 |
| ISP | Verizon Internet Services |
| Organization | Verizon Internet Services |
| Netspeed | Cable/DSL |
| Domain Name | verizon.net |

MaxMind, GeoIP and related marks are registered trademarks of MaxMind, Inc.
Copyright © 2010 MaxMind, Inc. All Rights Reserved. [Terms of use](#).

Lab 1 – IP Geolocation

- First browse to google.com, and search for “what is my ip address”
 - Record the IP Address.
 - Do the same thing with a different device (cell phone)
- Next, browse to
 - <http://hostip.info>
 - Enter the IP Address and press go.

What your browser exposes

- By Default
- By HTTP Cookies
- By Scripting
- By External Connections
- **Follow along and go to the various websites...**

By Default

- Your IP Address / host name
 - REMOTE_ADDR : 70.108.107.180
 - REMOTE_HOST : pool-70-108-107-180.res.east.verizon.net
- User Agent (what web browser you use)
 - HTTP_USER_AGENT : Mozilla/5.0 (Windows; U; Windows NT 6.0; en-US; rv:1.9.2.6) Gecko/20100625 Firefox/3.6.6 (.NET CLR 3.5.30729)
 - Note language is exposed.
- Cookies (if set)
 - HTTP_COOKIE : __utma=145017023.1934880434.1277898848.1278891119.1278893601.5; __utmz=145017023.1277898848.1.1.utmcsr=(direct)|utmccn=(direct)|utmcmd
 - This was set via a google analytics script I run to see who visits my main zak.freeshell.org page.
- <http://zak.freeshell.org/env.pl>

By HTTP Cookies

- Cookies were originally designed so that you could have a “stateful” web session for shopping.
- A server sets a key value pair when returning a webpage or image. You send that key value pair with every request to that domain, until it expires.
- Cookies only return information to the domain that you are visiting
 - Note, though, if you embed an image from a different server, that server can set a separate cookie. This is called a third party cookie.

By Scripting

- How to tell what company you're from
- How to tell if you've visited other sites (HREF color JavaScript tester)
- JavaScript Location API (Where you are)
- "Click Heat" – where you've clicked
- Ever Cookie.js – Have I seen you before?

How to tell what company you're from

- Via <http://panopticklick.eff.org/> (with many other attributes you can get from scripting)
- ▶ Flash code identifies system installed fonts
 - ▶ *Do you have a “company “ font? See if you can find it!*

What you've visited before script

- Found in the wild.
- ***Used to work.***
 - Last year Firefox / other browsers have been patched against this type of exploit.
- Uses JavaScript to look at `` html to see what the *color* of the link is , to see if it's been visited before
- Example: **Visited** | **Not Visited**
- See <http://www.techdirt.com/articles/20101130/21535012065/how-youporn-tries-to-hide-that-its-spying-your-browsing-history.shtml> or <http://preview.tinyurl.com/3yqbptk>

JavaScript Location API

- New as of 10 February 2010
- <http://dev.w3.org/geo/api/spec-source.html>
- Will it work? It's still a bit of a new feature, so it *should* on most browsers.
 - See <http://html5demos.com/> for html5 integration and <http://html5demos.com/geo> for a geo location example.
- Often uses Wi-Fi to detect location (unless a mobile device and it uses GPS)

Sample Location Code

```
<HTML><BODY>
<SCRIPT TYPE="text/JavaScript">
  function showMap(position) {
    alert("Lat:" + position.coords.latitude + "," + "Long:" + position.coords.
longitude);
  }

  // One-shot position request.
  if (navigator.geolocation){
    navigator.geolocation.getCurrentPosition(showMap);
  } else {
    alert('No geolocation available. ');
  }
</script>
</BODY></HTML>
```

Notes on Location

- Browser to OS for location information
 - Insert magic here in the OS level.
- Location returned is WGS84 standard.
- Can be off by a large distance
 - I'm sometime geolocated to Baltimore instead of Arlington.
 - When I was in Bedford, I was geolocated to south of Boston

“ClickHeat”

- <http://www.labsmedia.com/clickheat/index.html>
- Tracks where users click on your website via simple JavaScript.

“ClickHeat” Demo slide

The screenshot shows the ClickHeat website with a heatmap overlay. The heatmap consists of numerous blue dots of varying sizes, indicating click locations and intensity. The website layout includes a navigation menu at the top, a sidebar with filters, and a main content area with text and links.

Website & Group: clickheat

Browser: All

Heatmap and its transparency: [Slider]

Screen size: All

Calendar: October 1, 2, 3, 4, 5, 6, 7, 8, 9, 10, 11, 12, 13, 14, 15, 16, 17, 18, 19, 20, 21, 22, 23, 24, 25, 26, 27, 28, 29, 30, 31

ClickHeat | Clicks heatmap

ClickHeat is a visual heatmap of clicks on a HTML page, showing hot and cold click zones. ClickHeat is an OpenSource software, released under GPL licence, and free of charge.

Requirements

- on the browser's client: Javascript (tested on Firefox 2.0, Internet Explorer 6 and 7, Konqueror...)
- on the server: either Linux or Windows (since ClickHeat 1.3 release), Apache or Lighttpd (other may work fine), PHP, the graphic library GD2 (PNG support needed). Please post on the [bug tracker](#) or contact us (link at the bottom of the page) if you have problem running ClickHeat.

More information:

- [Installation and upgrade of ClickHeat](#)
- [Heatmap Class](#)
- [Frequently Asked Questions](#)
- [Thanks](#)
- [Performance and optimization](#)
- Ressources**
- [Download](#)
- [demo/demo](#)

labs media

ClickHeat is a visual heatmap of clicks on a HTML page, showing hot and cold click zones. ClickHeat is an OpenSource software, released under GPL licence, and free of charge.

Our projects

Features

Internet 100%

Evercookie.js

- Uses HTML 5 features ; flash ; JavaScript; and many other techniques to create an “evercookie” that can’t be revoked*.
- See <http://samy.pl/evercookie/>
- *Easily anyways.
 - <http://arstechnica.com/security/news/2010/10/it-is-possible-to-kill-the-evercookie.ars>
 - <http://preview.tinyurl.com/29ka7ba>

By External Connections

- Browsers allow for multiple types of connections
 - For example: http ; ftp ; Microsoft shares ; etc.
- By including a href link to a different type of connection, you may be bypassing proxy settings and thus exposing additional information about yourself.
 - Possibly your username / domain
 - Your real IP Address (because it's not proxied.)
- You can unknowingly request a different protocol by visiting a web page.
- See http://www.defcon.org/images/defcon-17/dc-17-presentations/defcon-17-gregory_fleischer-attacking_tor.pdf or <http://preview.tinyurl.com/brcg6ca>

External Example – what you see

Edit and Click Me >>

Your Result:

```
<html>
<body>
<IMG SRC="ftp://freeshell.org/etc/motd" width=0 height=0>
</body>
</html>
```

Edit the code above and click to see the result.

W3Schools.com - Try it yourself

This is common!

- This type of identification is profitable!
 - Multiple companies are interested in “fingerprinting” PC’s so they have a permanent record of your machine’s web accesses.
 - <http://online.wsj.com/article/SB10001424052748704679204575646704100959546.html> or <http://preview.tinyurl.com/32c9frf>

See also

- http://waxy.org/2011/11/google_analytics/
 - Google Analytics A Potential Threat to Anonymous Bloggers
- <https://github.com/michaelhans/derezzed-light>
 - Takes panopclick attributes via JavaScript, converts it to a browser key, and turns it into a logon mechanism for a website.

Questions?

- Let's have a break

When we get back, we start with ways of preventing these mechanisms from working...