ELECTRONIC FRONTIER FOUNDATION
DEFENDING YOUR RIGHTS IN THE DIGITAL WORLD

SEARCH

JANUARY 27, 2010 | BY PETER ECKERSLEY

Browser Versions Carry 10.5 Bits of Identifying Information on Average

This is part 3 of a series of posts on user tracking on the modern web. You can also read part 1 and part 2.

Whenever you visit a web page, your browser sends a "User Agent" header to the website saying precisely which operating system and web browser you are using. This information could help distinguish Internet users from one another because these versions differ, often considerably, from person to person. We recently ran an experiment to see to what extent this information could be used to track people (for instance, if someone deletes their browser cookies, would the User Agent, alone or in combination with some other detail, be unique enough to let a site recognize them and re-create their old cookie?).

Our experiment to date has shown that the browser User Agent string usually carries 5-15 bits of identifying information (about 10.5 bits on average). That means that on average, only one person in about 1,500 ($2^{10.5}$) will have the same User Agent as you. On its own, that isn't enough to recreate cookies and track people perfectly, but in combination with another detail like geolocation to a particular ZIP code or having an uncommon browser plugin installed, the User Agent string becomes a real privacy problem.

User Agents: An Example of Browser Characteristics Doubling As Tracking Tools

When we analyze the privacy of web users, we usually focus on user accounts, cookies, and IP addresses, because those are the usual means by which a request to a web server can be associated with other requests and/or linked back to an individual human being, computer, or local network.

Typical advice for improving your privacy as you surf the web might include blocking or deleting cookies (and supercookies), and using proxy servers or tools like Tor to hide your IP address.

It's not intuitively obvious that a User Agent poses a similar risk to a unique tracking cookie. After all, cookies were designed, in part, to help web sites distinguish and recognize individual browsers, and User Agents weren't. And there could be millions of people out there who use the same browser and operating system that you do. But let's examine the matter more closely. A typical User Agent string looks something like this:

```
Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.1.3) Gecko/20090824 Firefox/3.5.3 (.NET CLR 3.5.30729)
```

In fact, that was the most common user agent string among browsers visiting the EFF website during the test period: Firefox 3.5.3 running on Windows XP. Notice that the operating system and browser versions are extremely specific and that the User Agent also includes the user's preferred language. There are a lot of things that can vary inside that string, and those variations can be used to distinguish and track people as they browse the Web.

Our Results to date on User Agent Identifiability

We ran an experiment to measure precisely how identifying the User Agent strings would have been among a 36-hour anonymized sample of requests to the EFF website. The following table shows different classes of browser, with the number of bits for best and average case User Agents within that class:

Identifying information in various classes of browsers

Browser class	Avg. identifying information	Minimum identifying information	(Least identifying user agent)

Stay in Touch

Email Address

Postal Code (optional)

SIGN UP NOW

NSA Spying

eff.org/nsa-spying

EFF is leading the fight against the NSA's illegal mass surveillance program. Learn more about what the program is, how it works, and what you can do.

Follow EFF

Europe's new data protection law has a hidden flaw: a takedown process that's worse than the DMCA: https://www.eff.org/deeplinks...
NOV 20 @ 3:28PM

Everyone agrees: The "staggering concentration" of patent cases in just a few courts is bad for the patent system https://www.washingtonpost.co..
NOV 20 @ 3:18PM

Europe will soon adopt a powerful new data protection regulation, but it could do unintended harm to free expression https://www.eff.org/deeplinks...
NOV 20 @ 12:59PM

Twitter   Facebook   Identi.ca

Projects

Bloggers' Rights

Coders' Rights

Free Speech Weak Links

2015年11月22日 20:31

Modern Windows Desktops	10.3-11.3 bits	4.6 - 5.0 bits	Mozilla/5.0(Windows; U; Windows NT 5.1; en-US; rv:1.9.1.3) Gecko/20090824 Firefox/3.5.3 (.NET CLR 3.5.30729)
Internet Explorer	13.2-13.5 bits	6.3 - 7.2 bits	Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1)
Firefox	8.6 - 9.4 bits	4.6 - 5.0 bits	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US; rv:1.9.1.3) Gecko/20090824 Firefox/3.5.3 (.NET CLR 3.5.30729)
Chrome	7.5-8.5 bits	5.7 - 6.2 bits	Mozilla/5.0 (Windows; U; Windows NT 5.1; en-US) AppleWebKit/532.0 (KHTML, like Gecko) Chrome/3.0.195.27 Safari/532.0
Linux	11.8-13.15 bits	6.6-7.9 bits	Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.0.14) Gecko/2009090216 Ubuntu/9.04 (jaunty) Firefox/3.0.14
Ubuntu	9.6 - 11.7 bits	6.6 - 7.8 bits	Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.0.14) Gecko/2009090216 Ubuntu/9.04 (jaunty) Firefox/3.0.14
Debian	13.5-15.3 bits	10.50 - 11.7 bits	Mozilla/5.0 (X11; U; Linux i686; en-US; rv:1.9.0.14) Gecko/2009091010 Iceweasel/3.0.6 (Debian-3.0.6-3)
Macintosh	8.8-9.3 bits	5.8-5.8 bits	Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.5; en-US; rv:1.9.1.3) Gecko/20090824 Firefox/3.5.3
iPhone	10.8 - 11.3 bits	8.7 - 9.3 bits	Mozilla/5.0 (iPhone; U; CPU iPhone OS 3_1 like Mac OS X; en-us) AppleWebKit/528.18 (KHTML, like Gecko) Version/4.0 Mobile/7C144 Safari/528.16
Blackberry	14.7 - 15.5 bits	12.0 - 12.7 bits	BlackBerry9530/4.7.0.148 Profile/MIDP-2.0 Configuration/CLDC-1.1 VendorID/105
Android	14.4 - 14.4 bits	12.2-12.4 bits	Mozilla/5.0 (Linux; U; Android 1.6; en-us; T-Mobile G1 Build/DRC83) AppleWebKit/528.5+ (KHTML, like Gecko) Version/3.1.2 Mobile Safari/525.20.1

There are several remarkable facts about this dataset. Overall, it's amazing how identifying User Agent strings are. 10.5 bits is about one-third of the total information required to identify an Internet user.

It's also surprising that platforms like Firefox and Ubuntu, which have lower market penetration, are on average comparable or even less identifying than Windows and Microsoft Internet Explorer, which have very large userbases and should therefore have larger crowds to hide in. Part of this may be that visitors to the EFF website are over-representative of the former groups, but it's also clear that a large part of this is that Internet Explorer has a very high level of variation in its User Agent strings, with typical examples looking something like this:

Mozilla/4.0 (compatible; MSIE 8.0; Windows NT 6.0; Trident/4.0; SLCC1; .NET CLR 2.0.50727; Media Center PC 5.0; .NET CLR 3.5.30729; .NET CLR 3.0.30618)

All of the different library and component versions there essentially function as partial tracking tokens.

We've launched a project called Panopticlick to collect a new dataset that extends this analysis from User Agents to the full browser plugin and configuration space. You can use Panopticlick to receive a uniqueness measurement for your own browser, and help EFF's privacy research efforts at the same time!

Methodology

During September 2009, we took a 36 hour sample of anonymized requests to the eff.org web server by hashing the IP address of each request with a random salt, and throwing away the salt. We then calculated the amount of identifying information conveyed by each browser. Identifying information is measured in "bits of entropy", and says how large a crowd the information would reveal you within. Browsers usually convey between 5 and 15 bits of identifying information, about 10.5 bits on average. 10 bits of identifying information would allow you to be picked out of a crowd of 2^{10}, or 1024 people. 10.5 bits of information identifies can identify people from crowds of just under 1,448.

Because we did not use cookies or any other mechanism to distinguish between repeat and new visitors, each measurement of bits of identifying information lies between an upper and lower bound.[1]

---

1. One bound is based on a count in which each hashed IP address is counted for only one request; the other bound is based on treating each hit as a unique browser. In almost all cases, the true amount of identifying information pertaining to the *browser* should lie between these two values.

Privacy

MORE DEEPLINKS POSTS LIKE THIS   RECENT DEEPLINKS POSTS

JANUARY 2010

A Primer on Information Theory and Privacy

SEPTEMBER 2009

New Cookie Technologies: Harder to See and Remove, Widely Used to Track You

NOVEMBER 2012

Privacy in Ubuntu 12.10: Full Disk Encryption

MARCH 2013

Firefox's new, smarter cookie policy is a privacy win for users

MARCH 2009

Google Begins Behavioral Targeting Ad Program

NOV 20, 2015

EFF Joins Broad Coalition of Groups to Protest the TPP in Washington D.C.

NOV 20, 2015

Unintended Consequences, European-Style: How the New EU Data Protection Regulation will be Misused to Censor Speech

NOV 20, 2015

New Report Rates Peruvian ISPs: Who Defends Your Data?

NOV 19, 2015

Nuevo reporte muestra qué ISPs peruanas resguardan la privacidad de usuarios

NOV 19, 2015

YouTube Backs Its Users With New Fair Use Protection Program

DEEPLINKS TOPICS

Fair Use and Intellectual Property: Defending the Balance

Free Speech

Innovation

International

Know Your Rights

Privacy

Trade Agreements and Digital Rights

Security

State-Sponsored Malware

Abortion Reporting

Analog Hole

Anonymity

Anti-Counterfeiting Trade Agreement

Biometrics

Bloggers' Rights

Broadcast Flag

Broadcasting Treaty

CALEA

Cell Tracking

Coders' Rights Project

Computer Fraud And Abuse Act Reform

Content Blocking

Copyright Trolls

Council of Europe

Cyber Security Legislation

CyberSLAPP

Defend Your Right to Repair!

Development Agenda

Digital Books

Digital Radio

Digital Video

DMCA

DMCA Rulemaking

Do Not Track

DRM

E-Voting Rights

EFF Europe

Encrypting the Web

Export Controls

FAQs for Lodsys Targets

File Sharing

Fixing Copyright? The 2013-2015 Copyright Review Process

FTAA

Genetic Information Privacy

Hollywood v. DVD

How Patents Hinder Innovation (Graphic)

ICANN

International Privacy Standards

Internet Governance Forum

Law Enforcement Access

Legislative Solutions for Patent Reform

Locational Privacy

Mandatory Data Retention

Mandatory National IDs and Biometric Databases

Mass Surveillance Technologies

Medical Privacy

National Security and Medical Information

National Security Letters

Net Neutrality

No Downtime for Free Speech

NSA Spying

OECD

Offline : Imprisoned Bloggers and Technologists

Online Behavioral Tracking

Open Access

Open Wireless

Patent Busting Project

Patent Trolls

Patents

PATRIOT Act

Pen Trap

Policy Analysis

Printers

Public Health Reporting and Hospital Discharge Data

Reading Accessibility

Real ID

RFID

Search Engines

Search Incident to Arrest

Section 230 of the Communications Decency Act

Social Networks

SOPA/PIPA: Internet Blacklist Legislation

Student and Community Organizing

Stupid Patent of the Month

Surveillance and Human Rights

Surveillance Drones

Terms Of (Ab)Use

Test Your ISP

The "Six Strikes" Copyright Surveillance Machine

The Global Network Initiative

The Law and Medical Privacy

TPP's Copyright Trap

Trans-Pacific Partnership Agreement

Travel Screening

TRIPS

Trusted Computing

Video Games

Wikileaks

WIPO

Transparency

Uncategorized