

Lessons from privacy measurement

Arvind Narayanan

Princeton University

@random_walker

Caveat: my work is in the web privacy space

BUT I've aimed to extract broadly applicable lessons

Common theme: issues beyond encryption

Outline of this talk

- The ship has not sailed
- Privacy attitudes and technologies evolve rapidly; how can standards cope?
- Measurement: why it matters and how to preserve it



A research project of the **Electronic Frontier Foundation**

Panopticlick

How Unique – and Trackable – Is Your Browser?

Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

Panopticlick tests your browser to see how unique it is based on the **information** it will share with sites it visits. Click below and you will be given a uniqueness score, letting you see how easily identifiable you might be as you surf the web.

Only **anonymous data** will be collected by this site.



Panopticlick (2009)

Over 90% of users had a unique browser fingerprint

Fingerprinting is a privacy violation
Cannot be seen/controlled by user

AmlUnique (INRIA, France): similar conclusions

Learn how identifiable you are on the Internet



Help us investigate the diversity of web browsers.

This website aims at studying the diversity of browser fingerprints and providing developers with data to help them design good defenses. Contribute to the efforts by viewing your own browser fingerprint or consult the current statistics of data provided by users around the world!

[View my browser fingerprint](#)

Partial list of fingerprinting vectors

- User agent
- Accept header
- Content encoding
- Content language
- List of plugins
- Cookies enabled?
- Local/session storage?
- Timezone
- Screen resolution/depth
- List of fonts
- List of HTTP headers
- Platform
- Do Not Track
- Canvas
- WebGL
- Use of ad blocker

Conclusion: the horse has left the barn

Fingerprinting is devastatingly effective

Too late for anti-fingerprinting

(Me, until a year ago)

But wait...

users in previous studies self selected

New study:

- Only a third of users unique
- Mobile users: less than a fifth
- Number going down as Flash and Java phased out

Avoid privacy defeatism

The ship has not sailed

Imperfect defenses are still useful

Technology doesn't have to bear the full burden

Outline of this talk

- The ship has not sailed
- Privacy attitudes and technologies evolve rapidly; how can standards cope?
- Measurement: why it matters and how to preserve it

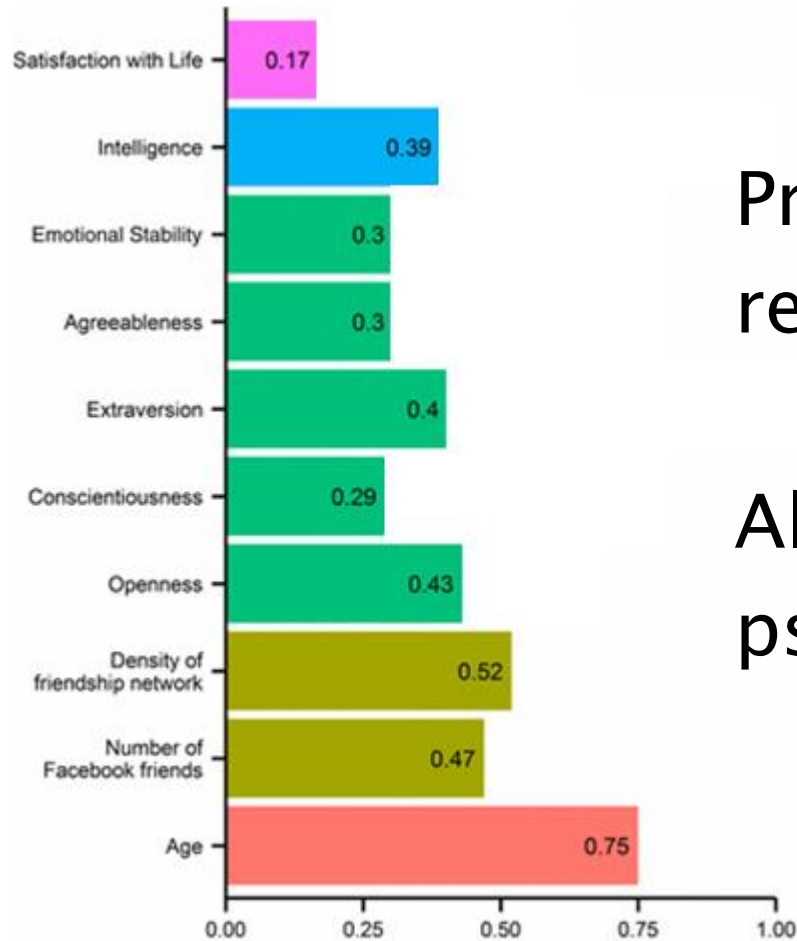
Privacy attitudes evolve quickly

Example: individual vs collective harms

Example: tradeoffs between privacy and other values

Result: Fixed technical definitions have difficulty capturing evolving norms and attitudes

Predicting sensitive traits from public FB “Likes”



Predicting “big 5” personality traits based on regression analysis of FB likes

Allegedly used by Cambridge Analytica for psychographic targeting

Kosinski et al: *Private traits and attributes are predictable from digital records of human behavior.* PNAS 2013.

How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

[+ Comment Now](#) [+ Follow Comments](#)

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. [Target](#), for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.



TARGET

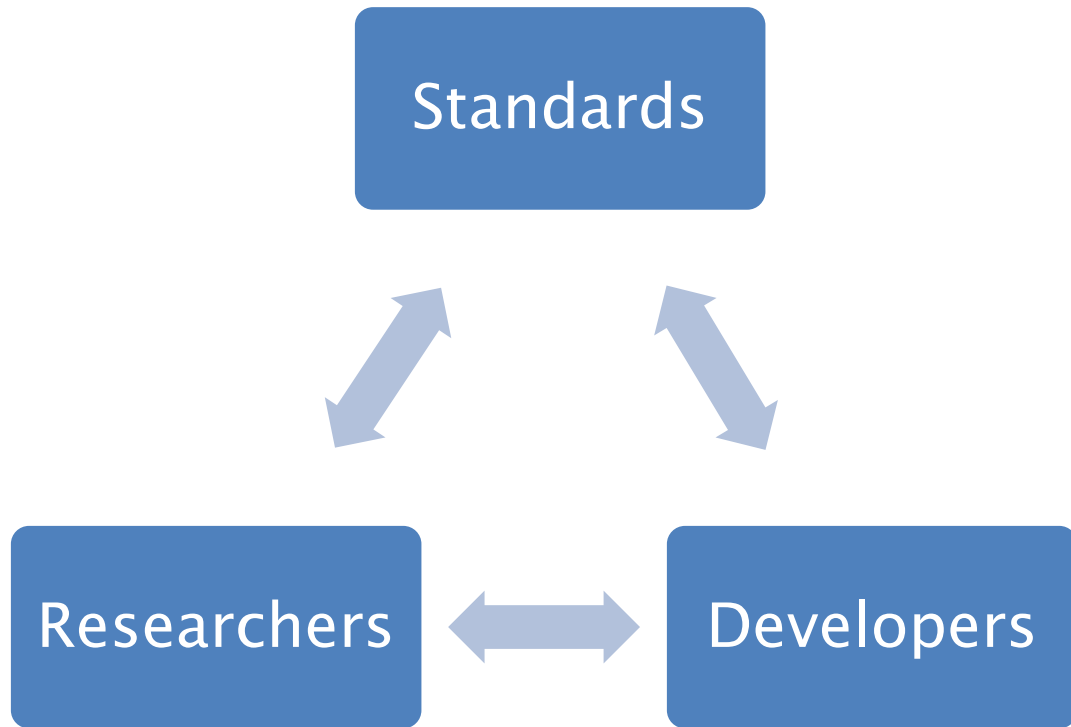
Target has got you in its aim

Privacy-infringing technologies evolve quickly

Paul Ohm's "database of ruin":

a single, massive database containing secrets about every individual, formed by linking different companies' data stores

Proposal: a tighter feedback loop



Incentivize academic researchers to

- Do privacy reviews of standards
- Study API use in the wild

Be explicit about assumptions

- Intended and unintended uses
- “Defense in depth” in case of misuse

Outline of this talk

- The ship has not sailed
- Privacy attitudes and technologies evolve rapidly; how can standards cope?
- **Measurement: why it matters and how to preserve it**

Measurement and privacy

Claim: measurement research has played a key role in keeping web privacy abuses in check

A tool for finding privacy violations

The screenshot shows the GitHub repository page for mozilla/OpenWPM. The repository is described as "A web privacy measurement framework" with a link to <https://webtap.princeton.edu/>. It has 1,291 commits, 7 branches, 20 releases, and 25 contributors. The page includes navigation tabs for Code, Issues (111), Pull requests (11), Projects (2), Wiki, Security, and Insights. A commit history table is visible at the bottom.

Commit	Message	Time
englehardt Merge pull request #410 from dashed/GH-408	Merge pull request #410 from dashed/GH-408	9 hours ago
automation	Merge pull request #410 from dashed/GH-408	9 hours ago
test	Add test page	20 days ago
.deploy-to-dockerhub.sh	Corrected the docker tag to push	4 days ago
.dockignore	Don't include .git in Docker images	8 days ago

Online Tracking: A 1-million-site Measurement and Analysis

Steven Englehardt
Princeton University
ste@cs.princeton.edu

Arvind Narayanan
Princeton University
arvindn@cs.princeton.edu

ABSTRACT

We present the largest and most detailed measurement of online tracking conducted to date, based on a crawl of the top 1 million websites. We make 15 types of measurements on each site, including stateful (cookie-based) and stateless (fingerprinting-based) tracking, the effect of browser privacy tools, and the exchange of tracking data between different sites (“cookie syncing”). Our findings include multiple sophisticated fingerprinting techniques never before measured in the wild.

This measurement is made possible by our open-source web privacy measurement tool, OpenWPM¹, which uses an

to resort to a stripped-down browser [31] (a limitation we explore in detail in Section 3.3). (2) We provide comprehensive instrumentation by expanding on the rich browser extension instrumentation of FourthParty [33], without requiring the researcher to write their own automation code. (3) We reduce duplication of work by providing a modular architecture to enable code re-use between studies.


Solving these problems is hard because the web is not designed for automation or instrumentation. Selenium,² the main tool for automated browsing through a full-fledged browser, is intended for developers to test their *own* websites. As a result it performs poorly on websites not con-

Impacts of web privacy measurement

- Enhancing blocklists
- Informing the public
- Correcting information asymmetry
- Convincing browser vendors to act
- Enforcement action in most egregious cases
- Informing policy makers

What about IoT?

 Most devices are end-to-end encrypted

 The two ends are the device and the server, not the user (or researcher)

⇒ Meaningful privacy measurement infeasible

The House That Spied on Me



If our smart lightbulbs are transmitting conversations
from our homes, do we have a way to know?

Google Calls Hidden Microphone in Its Nest Home Security Devices an 'Error'

The company says its was an oversight, but it does little to stem paranoia.



By [Sam Blum](#) Feb 21, 2019

Proposal: a debug mode for IoT devices

or researcher

When enabled, device allows user to intercept plaintext

Details and UX will depend on device

No technical way to prevent cheating;
reputational and legal incentives instead

Summary of this talk

- The ship has not sailed
- Privacy attitudes and technologies evolve rapidly; how can standards cope?
- Measurement: why it matters and how to preserve it