Reverse-engineering online tracking for privacy, transparency, accountability

webtap.princeton.edu

Arvind Narayanan
Princeton

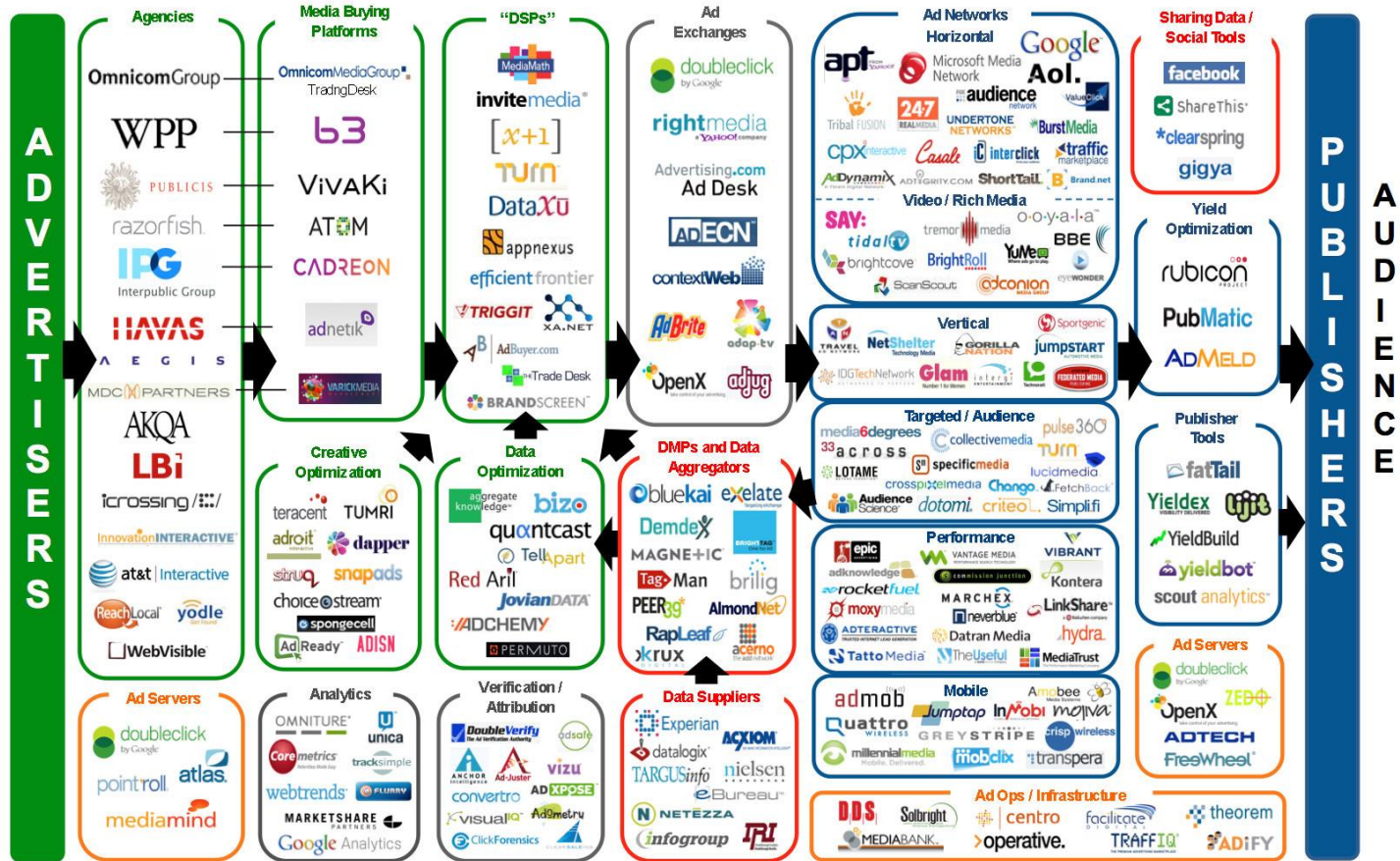Joint work with Steven Englehardt, Peter Zimmerman, Christian Eubank, …

"Third party" online tracking

Sites <u>other than</u> the one you're visiting

<u>typically invisible</u>

compiling <u>profiles</u> of your browsing history

64

Independent tracking mechanisms
on typical top-50 sites

How cookies can be used for global surveillance

DECEMBER 19, 2014 BY STEVEN ENGLEHARDT   LEAVE A COMMENT (EDIT)

Today we present an updated version of **our paper** examining how the ubiquitous use of online tracking cookies can allow an adversary conducting network surveillance to target a user or surveil users en masse. In the **initial version** of the study, summarized below, we examined the technical feasibility of the attack. Now we've made the attack model more complete and nuanced as well as analyzed the effectiveness of several browser privacy tools in preventing the attack. Finally, inspired by Jonathan Mayer and Ed Felten's **The Web is Flat** study, we incorporate the geographic topology of the Internet into our measurements of simulated web traffic and our adversary model, providing a more realistic view of how effective this attack is in practice.

From: IP 1.1.1.1
Cookie: "ID-A=aaa"

www.exampleA.com

Freedom to Tinker is hosted by Princeton's **Center for Information Technology Policy**, a research center that studies digital technologies in public life. Here you'll find comment and analysis from the digital frontier, written by the Center's faculty, students, and friends.
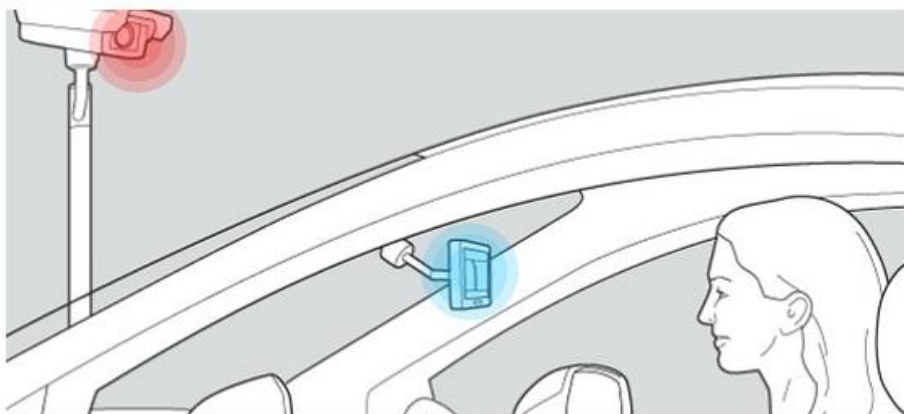

CENTER FOR INFORMATION TECHNOLOGY POLICY

Search this website...   **Search**

THE WALL STREET JOURNAL.

$12 *for* 12 Weeks   SUBSCRIBE NOW

U.S. EDITION ▼   Monday, September 29, 2014 As of 5:52 PM EDT

Subscribe | Log In

Home | World ▾ | U.S. ▾ | Business ▾ | **Tech ▾** | Markets ▾ | Market Data | Your Money ▾ | Opinion ▾ | Life & Culture ▾ | N.Y. ▾ | Real Estate ▾ | Management ▾

Digits | Personal Technology | **What They Know** | All Things Digital | CIO Journal

What They Know

2012: WATCHED | 2011: THE END OF PRIVACY | 2010: WHAT THEY KNOW

New Tracking Frontier: Your License Plates

The rise of license-plate tracking is a case study in how storing and studying people's everyday activities has become the default rather than the exception.

• ⧉ Interactive: The Surveillance Economy | ◥ Video: On the Prowl | Digits: Check Your Plate

Watched: A Wall Street Journal Privacy Report

As surveillance technologies decline in cost and grow in sophistication, tracking of many aspects of our daily activities, even the seemingly mundane, has become the default rather than the exception. The Wall Street Journal's Watched project—the latest in a years-long series on privacy—explores the impact of ubiquitous surveillance on citizens and society.

Multimedia

◀ 1 / 4 ▶
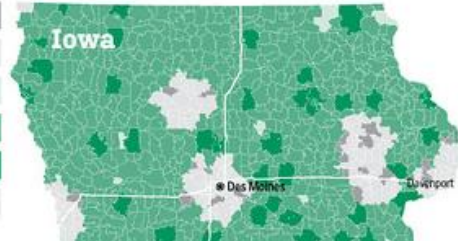
Likelihood of receiving higher prices, by ZIP code

Very low
Low
Middle
High
Very high
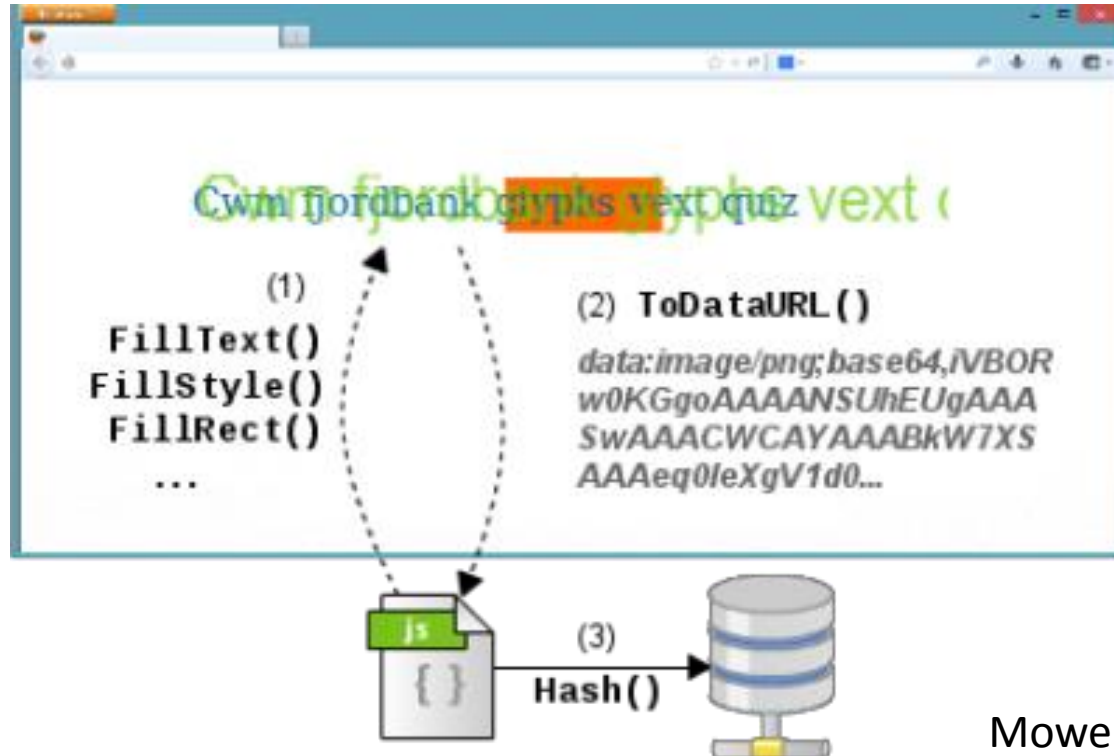No data

Iowa

● Des Moines   Davenport

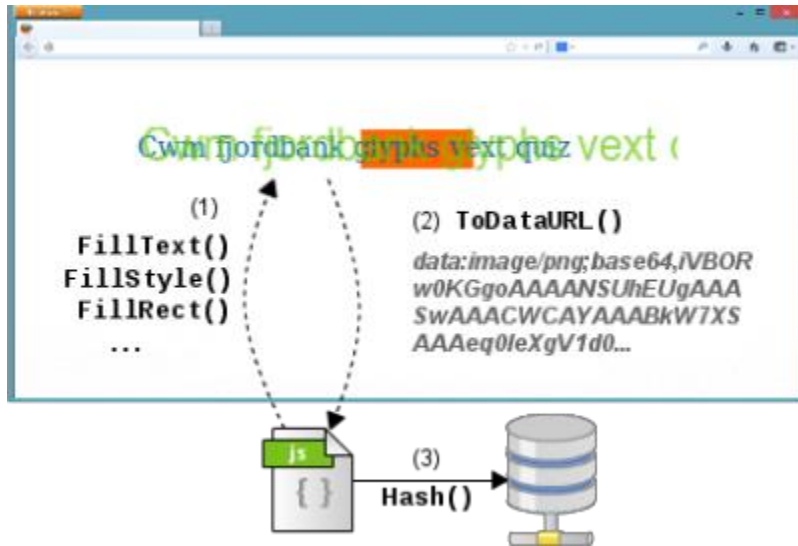# The unreasonable effectiveness of measurement

- Fixes "information asymmetry"

- More informed public debate

- Regulation and enforcement

What if oversight of
online tracking and personalization
could be *automated*?

Canvas fingerprinting

Mowery & Shacham '12

Canvas fingerprinting

100,000 site study

(KU Leuven & Princeton)

Used on over 5,500 sites

20 different providers

http://webtap.princeton.edu

## Princeton Web Transparency & Accountability Project

Measure Threats

We monitor websites and services to find out what user data companies collect, how they collect it, and what they do with it. With our measurement platform, we study privacy, security, and ethics of consumer data usage.

Create Change

Our external oversight exposes the privacy practices of companies and forces them to make improvements. In addition, the data and studies that we produce assist regulators and privacy tool developers in their efforts.

Inform the Public

We translate our research into practical information for public consumption. We aim to improve the accuracy of media reports about online privacy and to provide useful advice for consumers on this website.
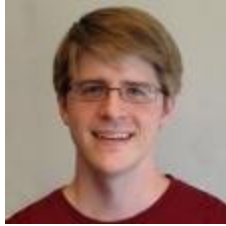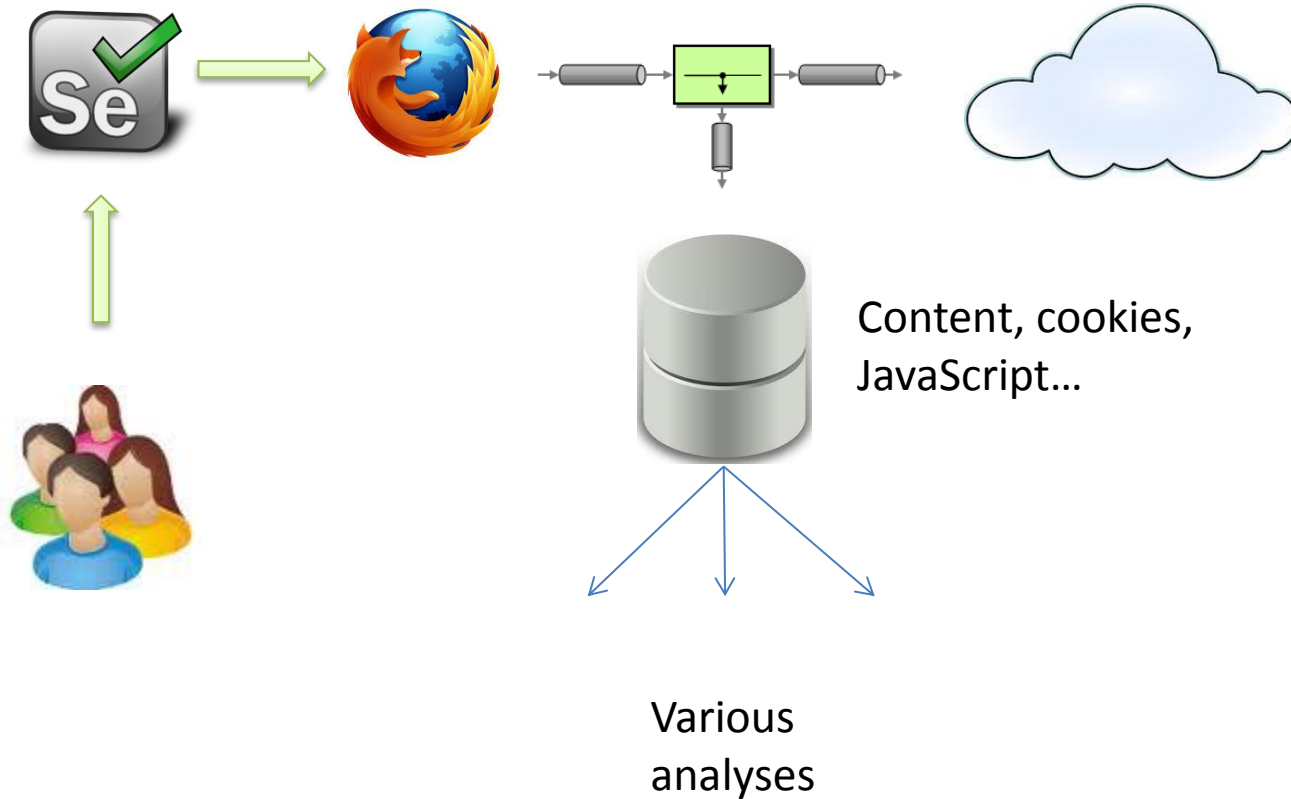
Steven Englehardt   Peter Zimmerman   Chris Eubank

OpenWPM architecture

Content, cookies, JavaScript…

Various analyses

Third-party tracking today is like spam 10 years ago

Solution: large datasets + machine learning

Beyond the web:
reverse-engineering algorithmic society

How can we have effective oversight of algorithmic personalization?

How to we deal with the vanishing boundary between online and off-line data collection?