

Reverse-engineering Online Tracking

From niche research field to easy-to-use tool

Steven Englehardt



webtap.princeton.edu



The New York Times - Breaking News, World News & Multimedia

The New York Times - Breaking... http://www.nytimes.com/

GUCCI The New York Times see what's inside

Advertising Advertising

Most Popular Times Topics Get Home Delivery-Bay Area Log In Register Times People

Tuesday, February 1, 2011 Last Update: 10:16 PM ET

Search Selective Foreign Advertising Follow Us Subscribe to Home Delivery

Switch to Global Edition

JOBS REAL ESTATE AUTOS ALL CLASSES

WORLD U.S. POLITICS NEW YORK BUSINESS DEALBOOK TECHNOLOGY SPORTS SCIENCE HEALTH OPINION ARTS Books Music Television Theater STYLE Dining & Wine Fashion & Style Home & Garden Weddings Celebrations TRAVEL All Blogs Caroons Classifieds Connections Crossword / Games Education First Look Learning Network Multimedia NYC Guide Obituaries Podcasts Public Editor Sunday Magazine T Magazine Video

Mubarak Takes New Term
Opposition Demands That He Leave Sooner

OPINION
EGYPT
Kristof: The White House should support the protesters in Egypt.
Blog: Mubarak's Speech
Follow on Twitter
Brooks: In the democratic...
Kobers: From a culture of...
Senator John Kerry: A...
MORE IN OPINION
Editorial: Debt Limit
Bloggingheads: Abortions

Log In With Facebook
Log in to see what your friends are sharing on nytimes.com.
Privacy Policy | What's This?

WHAT'S POPULAR NOW
The Paradox of Corporate Taxes in America
Why Not Regulate...
Social

MARKETS
As of 10:14 PM ET
JAPAN Nikkei 15,470.50
HongKong HangSeng 23,826.39
China CSI300 2,788.85
+198.00 +343.44 Closed
+1.91% +1.48% NYSE
NEW FELD SAVINGS ACCOUNT
1.30%
Advertising

GET QUOTES My Portfolios
Stock, ETFs, Funds Go

Video
The Mubarak...
Path to Change in Power Still Unclear
By ANTHONY SHADID 58 minutes ago
President Hosni Mubarak's vow to step down in the fall was not enough for the hundreds of thousands who poured into Tahrir Square.
Post a Comment | Read (332)

Quiet Acts of Protest on a Noisy Day
By KAREEM FAHMI and ANTHONY SHADID 54 minutes ago
Hundreds of thousands of Egyptians traveled like pilgrims to speak freely and to be heard.
King of Jordan Dismisses His Cabinet
Antiquities Chief Says Sites Are Largely Secure
New Service Lets Voices From Egypt Be Heard
The Lede: Updates From Day 8

INTERACTIVE FEATURE: A Timeline of Mubarak's Presidency
A chronology of President Hosni Mubarak's 30-year rule in Egypt.

More Photographs Interactive Map of Protests

iMeet built a meeting room just for you.
30-DAY FREE TRIAL

Source: Mayer & Mitchell; Third-Party Web Tracking: Policy and Technology

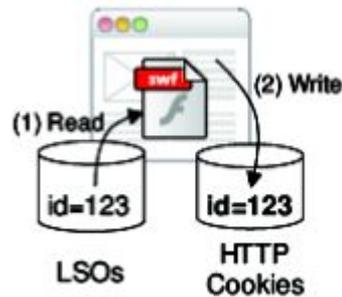
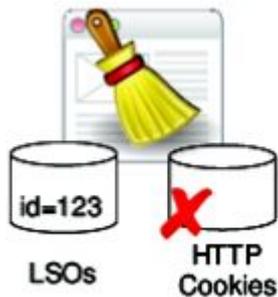
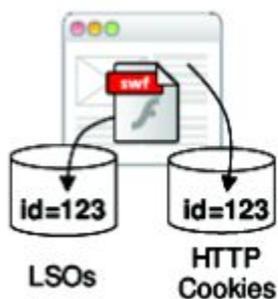


DELETE COOKIES?!

gegen-den-strich.com

FERNANDEZ

Evercookies



Respawn cookies using alternative locations

- Flash cookies, HTML5 localStorage, ETags, etc.

If you're going to track me, please use cookies

Ed Felten

July 7th, 2009

freedom-to-tinker.com



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

Panopti**cl**ick

How Unique — and Trackable — Is Your Browser?

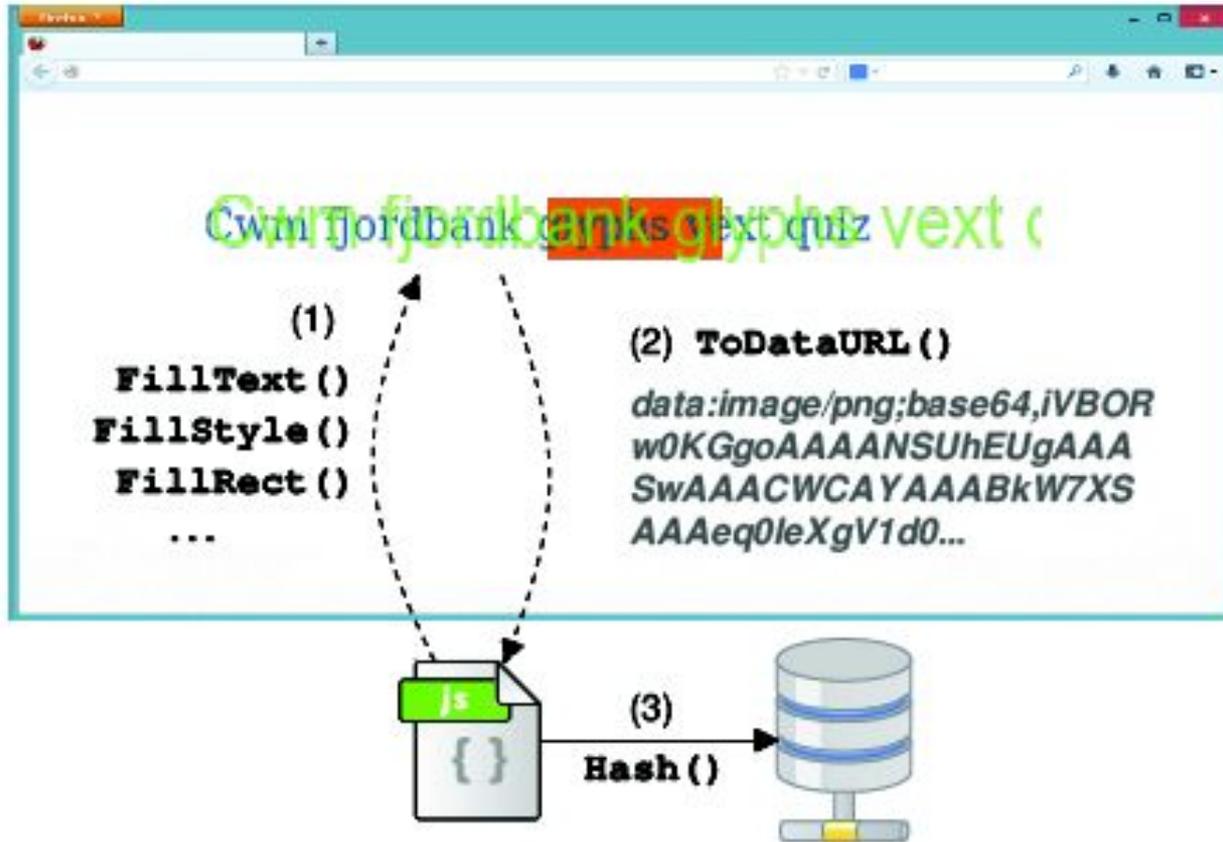
Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

Panopti**cl**ick 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.



Canvas Fingerprinting



2009

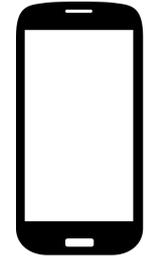
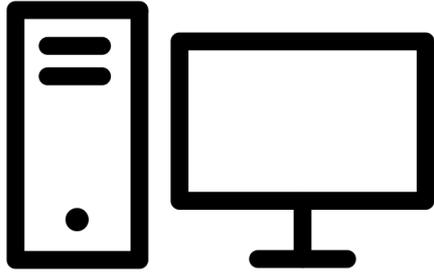
*If you're going to track me, please **use cookies***

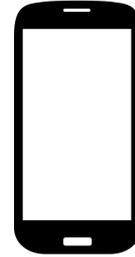
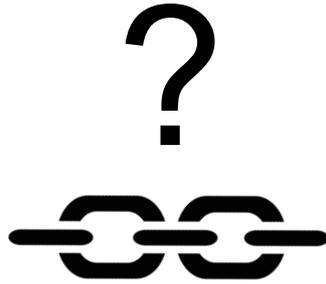
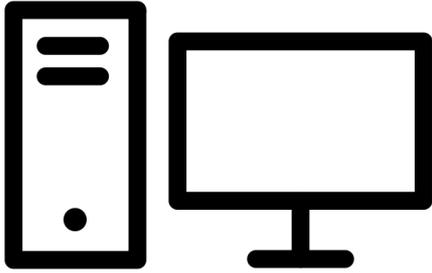
2009

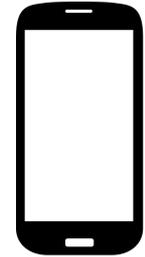
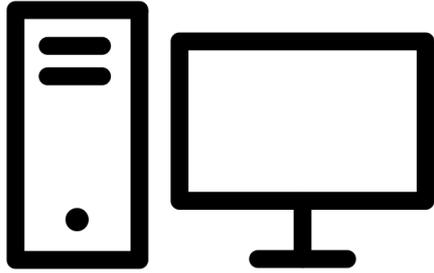
~~*If you're going to track me, please **use cookies***~~

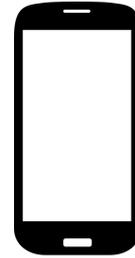
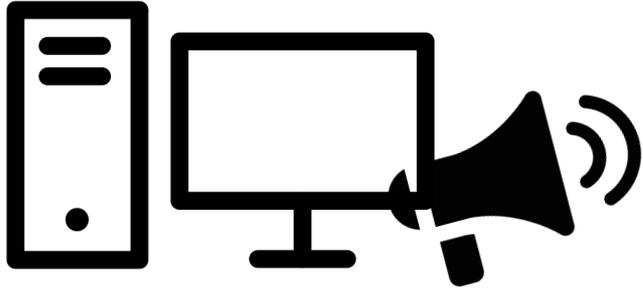
2010

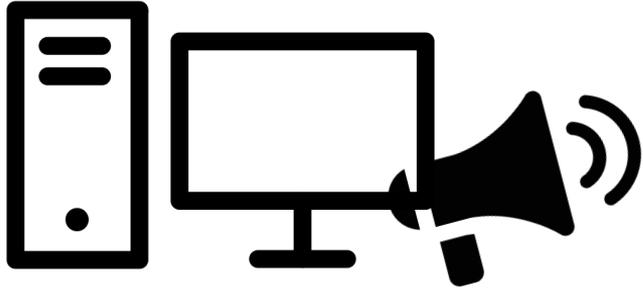
*If you're going to track me, please **use browser storage***













[News & Events](#) » [Events Calendar](#) » [Cross-Device Tracking](#)

Cross-Device Tracking



Re
Ma
FT
Tra
No
FT
for
Wc

NOV 16, 2015

CONSTITUTION CENTER

400 7th St SW, Washington, DC 20024 | [Directions & Nearby](#)

TAGS: [Bureau of Consumer Protection](#) | [Consumer Protection](#) | [Advertising and Marketing](#) |

2009

~~*If you're going to track me, please **use cookies***~~

2010

*If you're going to track me, please **use browser storage***

2009

~~If you're going to track me, please **use cookies**~~

2010

~~If you're going to track me, please **use browser storage**~~

2015

If you're going to track me, please **limit it to one device**

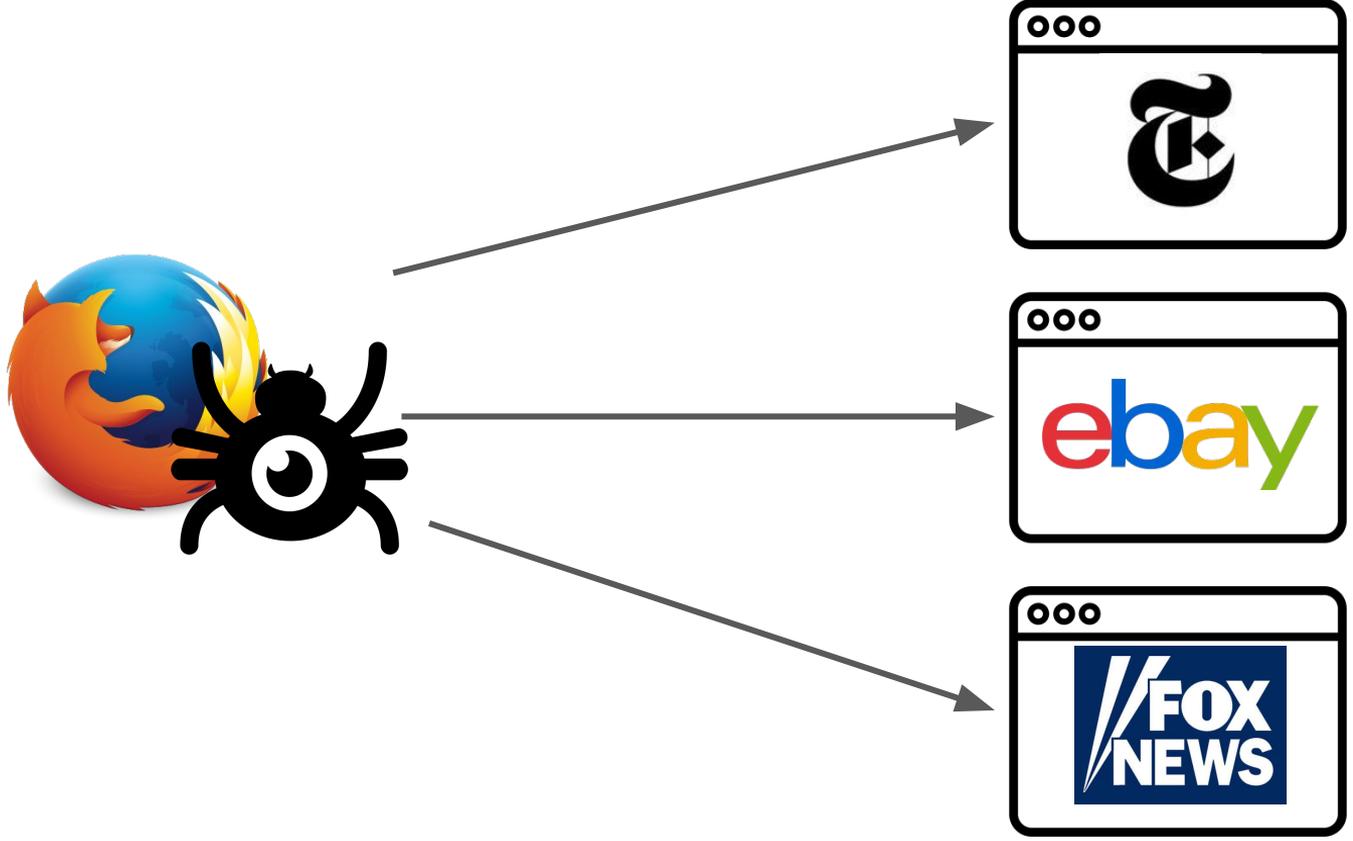
2015

~~If you're going to track me, please **limit it to one device**~~

2020

If you're going to track me, please _____?

Measurement can help!



Web measurement hurdles

1. Engineering Debt

Paper	Targets	Infrastructure		Variable				Scale		
		Automation ^{FF}	Instrumentation	Crowd-sourced Distributed	Location	User-agent	Demographics		Interests	Privacy Tools
Leakage of PII via OSN ('09) 31	PII leaks	M*	LHH							
Privacy diffusion on the web ('09) 30	Tracking: cookies	F,PS	Proxy					1.2K sites		
Challenges in measuring ('10) 25	Personalization: ads		Proxy			•	•	730 queries		
Flash cookies and privacy ('10) 53	Tracking: cookies, LSOs	M*						100 sites		
Privacy leakage in mOSN ('10) 32	PII leaks	M*	Proxy							
Flash cookies and privacy II ('11) 10	Tracking: cookies, LSOs	M*						100 sites		
Privacy leakage vs. protection measures ('11) 29	PII leaks	M*	Proxy					10 sites		
Respawn HTTP Cookies ('11) 41	Tracking: cookies, LSOs	UA*				•		600 sites		
Self-help tools ('11) 38	Tracking: cookies	UA*	FourthParty					500 sites		
Where everybody knows your username ('11) 39	PII leaks	M*	FourthParty				•	185 sites		
Detecting and defending ('12) 52	Tracking: cookies	FF, TT	TrackingTracker					2K sites		
Detecting price and search discrimination ('12) 42	Price discrimination	SA, CH, IE, JS	Proxy			•	•	•	200 sites	
Mac users steered to pricier hotels ('12) 37	Personalization: steering						•			
Measuring the effectiveness of privacy tools ('12) 11	Personalization: ads	F, SL						•		
Websites vary prices ('12) 57	Personalization: prices, deals						•			
What they do with what they know ('12) 60	Personalization: ads		Proxy					10 days		
AdReveal ('13) 34	Personalization: ads		Proxy, Ghostery				•	103K sites		
Cookieless monster ('13) 47	Tracking: fingerprinting							10K sites		
Crowd-assisted search ('13) 43	Price discrimination	F, CH	Custom plugin			•	•	•	600 sites	
Discrimination in online ad delivery ('13) 54	Ads	M, UA					•	•	2184 names	
FPDetective ('13) 7	Tracking: fingerprinting, JS	CR, SL, CJ, PJ	Proxy, Browser Code						1M sites	
Know your personalization ('13) 35	Personalization: search		Custom plugin			•			5K queries	
Measuring personalization of web search ('13) 26	Personalization: search	PJ				•	•		120 queries	
Who knows what about me? ('13) 36	PII leaks	F, PS, SL				•	•	•	1.5K sites	
Selling off privacy at auction ('13) 49	Cookie sync, bid prices	F, SL				•	•	•	5K sites	
Shining the floodlights ('13) 19	Tracking: cookies, JS	F, JS	FourthParty				•		500 sites	
Statistical approach ('13) 22	General tracking	F, PY	FourthParty					•	2K sites	
Adscape ('14) 13	Personalization: ads	F, SL	Custom plugin					•	10K sites	
Bobble ('14) 61	Personalization: search	CH, SL	Custom plugin			•	•	•	1K queries	
Information flow experiments ('14) 56	Personalization: ads	F, SL	Proxy					•		
Third-party OSN applications ('14) 14	PII leaks	F, SL	FourthParty					•	997 apps	
Price discrimination and steering ('14) 27	Price disc, steering	PJ				•	•	•	•	16 sites
Price discrimination of airline tickets ('14) 59	Price discrimination	CJ				•	•	•	•	21 days

^{FF} = Firefox, CH = Chrome, CR = Chromium, IE = Internet Explorer, SA = Safari, SL = Selenium, JS = JavaScript, PJ = PhantomJS, PS = PageStats, PY = Python, TT = TrackingTracker, CJ = CasperJS, UA = Unknown automation, M = manual, LHH = Live HTTP Headers, Asterisk = inferred

Many Studies, Many Platforms

- Automation:
 - 7 used Selenium (Full browser)
 - 4 used PhantomJS/CasperJS (Headless webkit)
- Instrumentation
 - 5 used FourthParty
 - 9 used a Proxy

Many Studies, Many Platforms

- Automation:
 - 7 used Selenium (Full browser)
 - 4 used PhantomJS/CasperJS (Headless webkit)
- Instrumentation
 - 5 used FourthParty
 - 9 used a Proxy

FourthParty is the only shared code

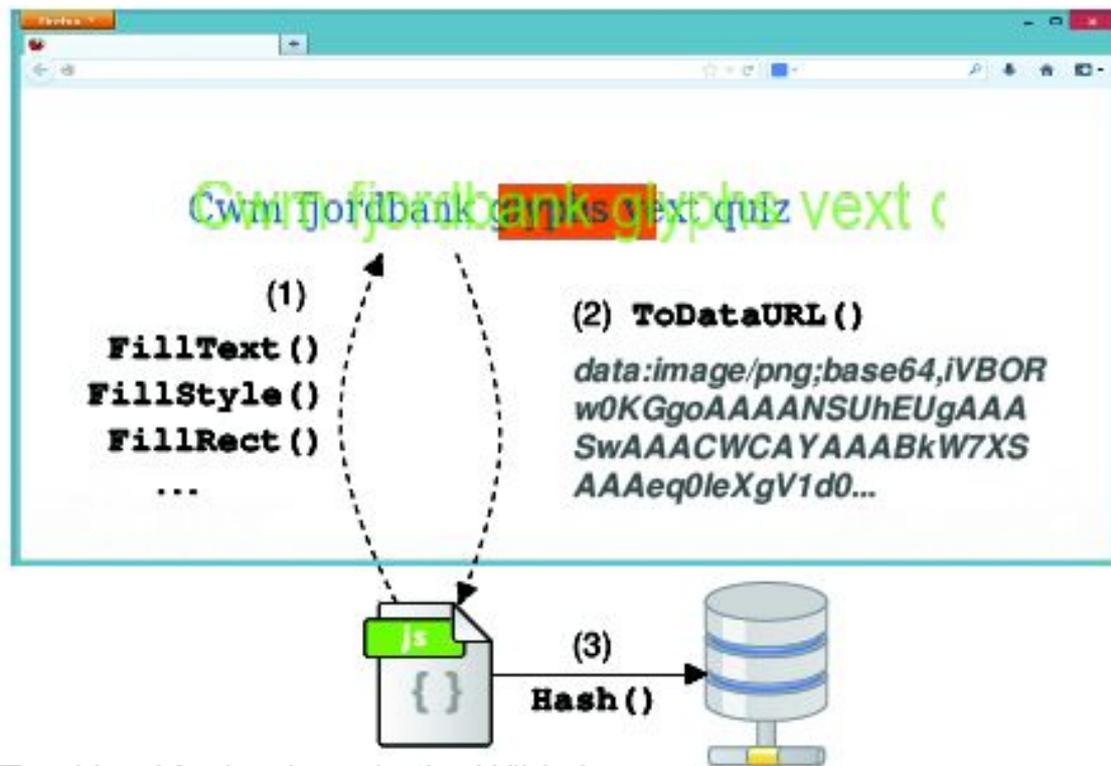
Web measurement hurdles

1. Engineering Debt

2. Lasting Impact

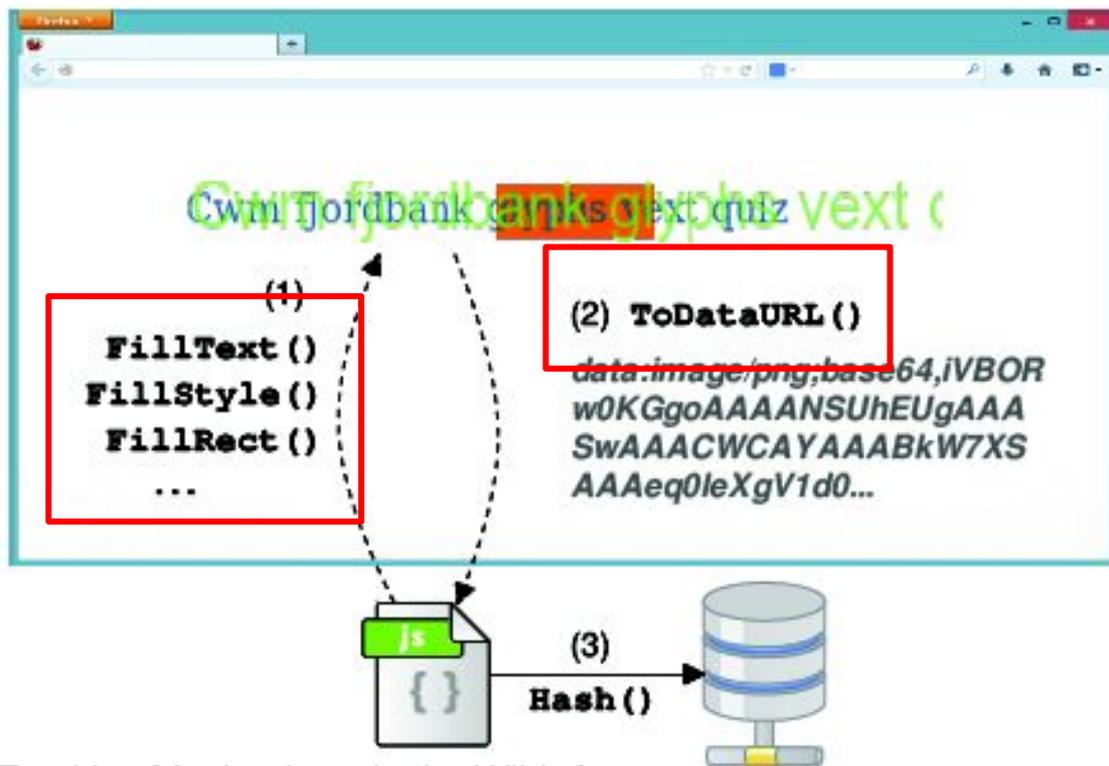
Canvas Fingerprinting in May 2014

- Acar, et.al (2014)
- 5% of Top 100k



Canvas Fingerprinting in May 2014

- Acar, et.al (2014)
- 5% of Top 100k



Canvas Fingerprinting in October 2015

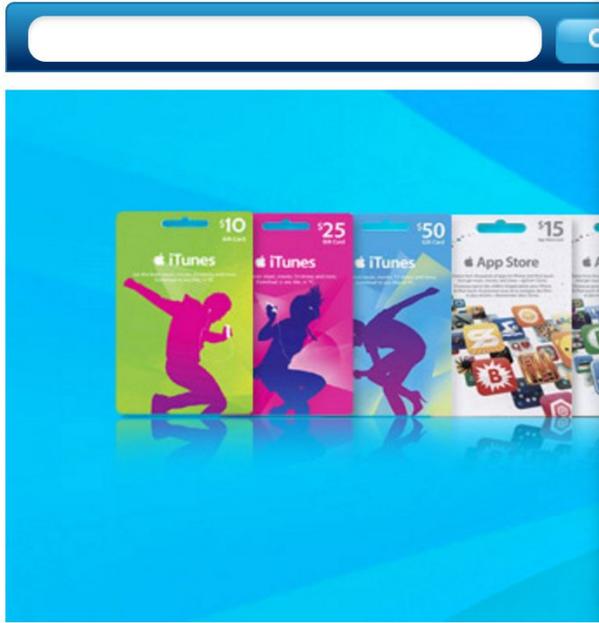
Over 100 first-party domains on the Top 100k



Canvas

Digital Delivery in Seconds!

Game Cards | Gift Cards | PrePaid Cards



Over

The W

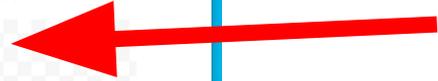


CanvasFingerprintBlock

Blocked 2 potential HTML canvas fingerprinting attempts on this page

Prevented a script on <http://www.pcgamesupply.com> from capturing the following 300px x 150px canvas:

<http://www.pcgamesupply.com>



Prevented a script on <http://www.pcgamesupply.com> from capturing the following 300px x 100px canvas:

@Browsers~%fingGPRint\$&, <canvas>

@Browsers~%fingGPRint\$&, <canvas>



Donate

15

100k

Overcoming these hurdles:

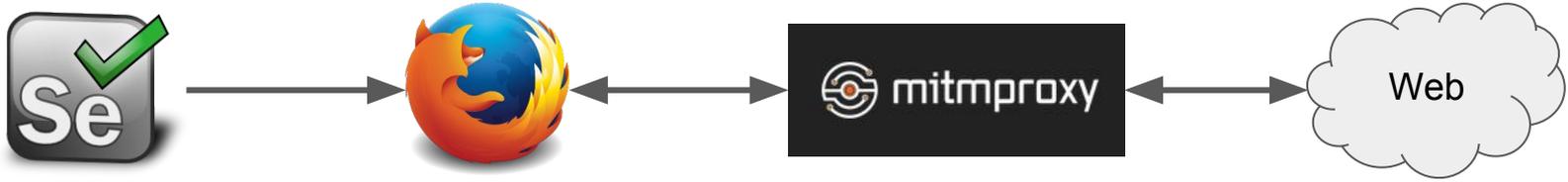
1. A Common Platform

2. A Web Privacy Census

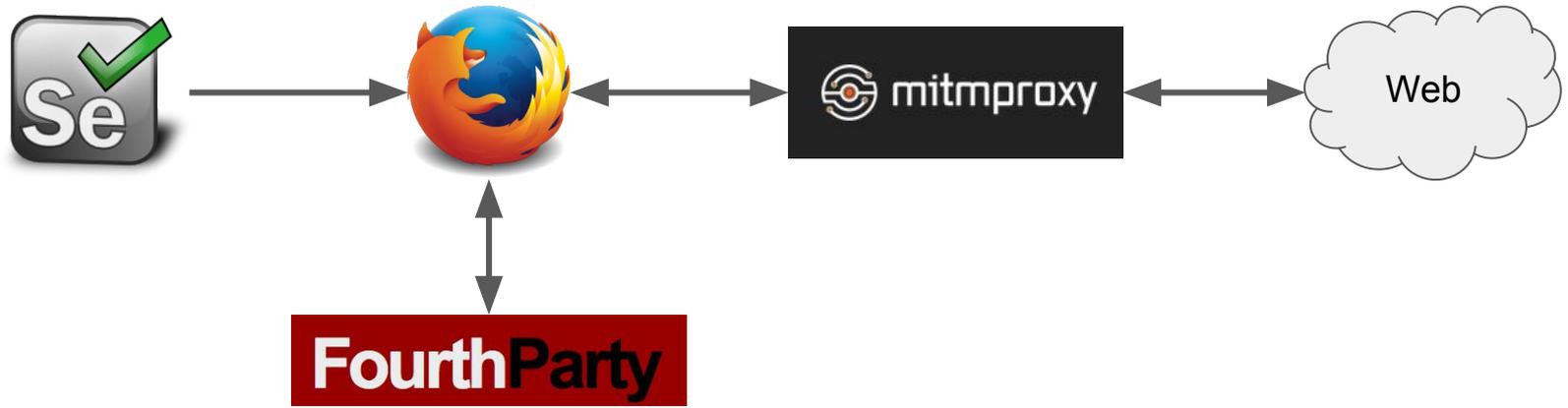
OpenWPM



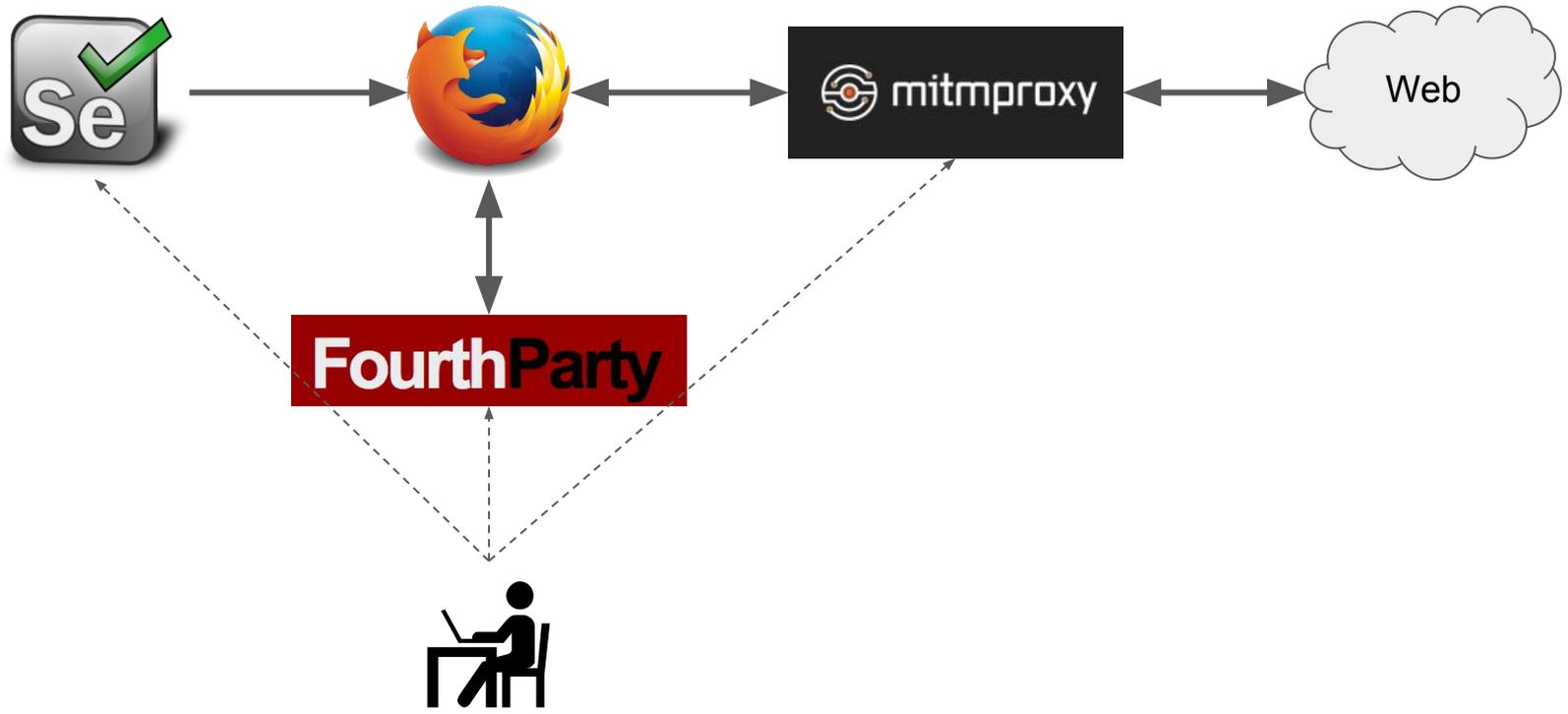
OpenWPM



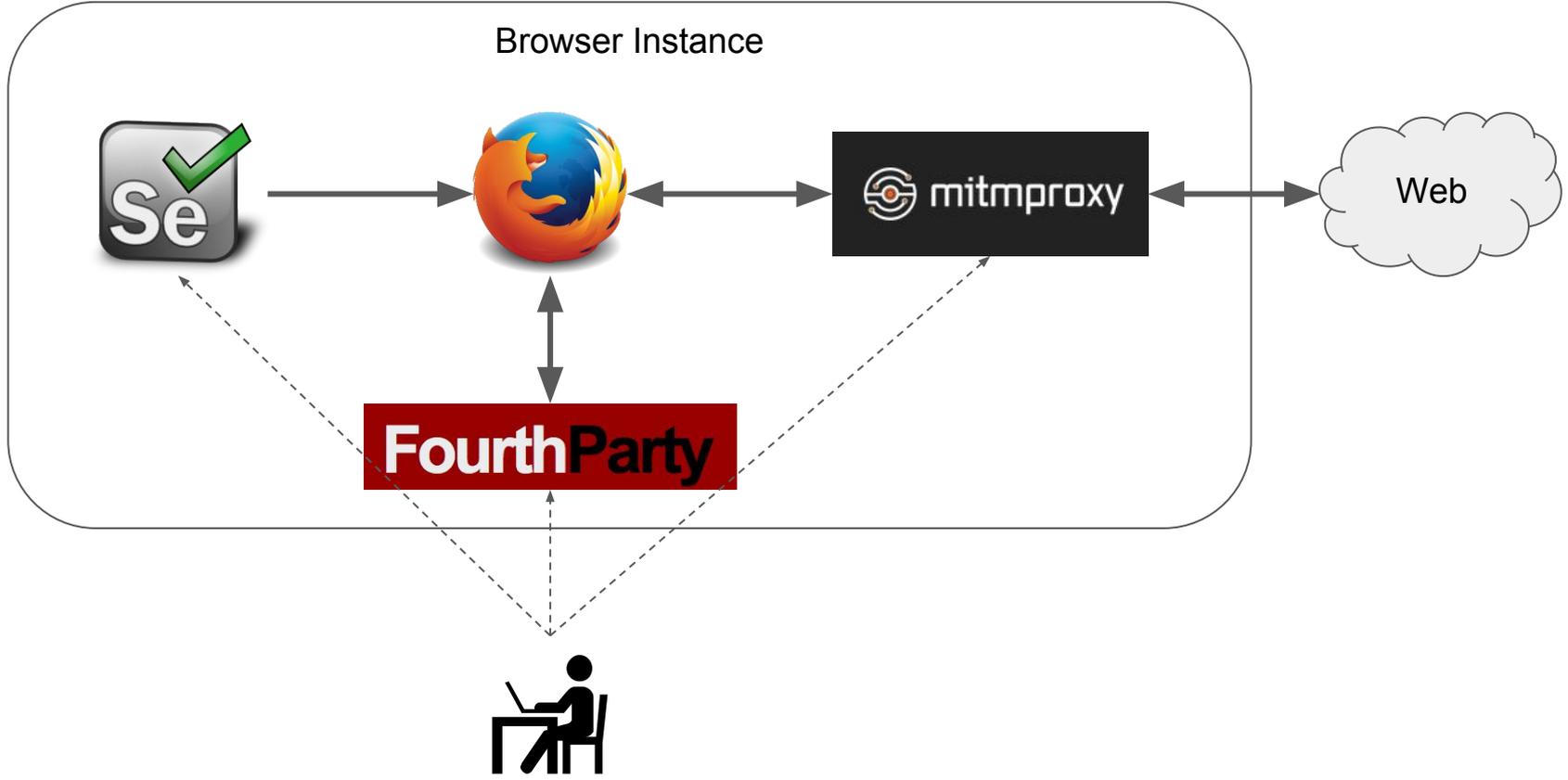
OpenWPM

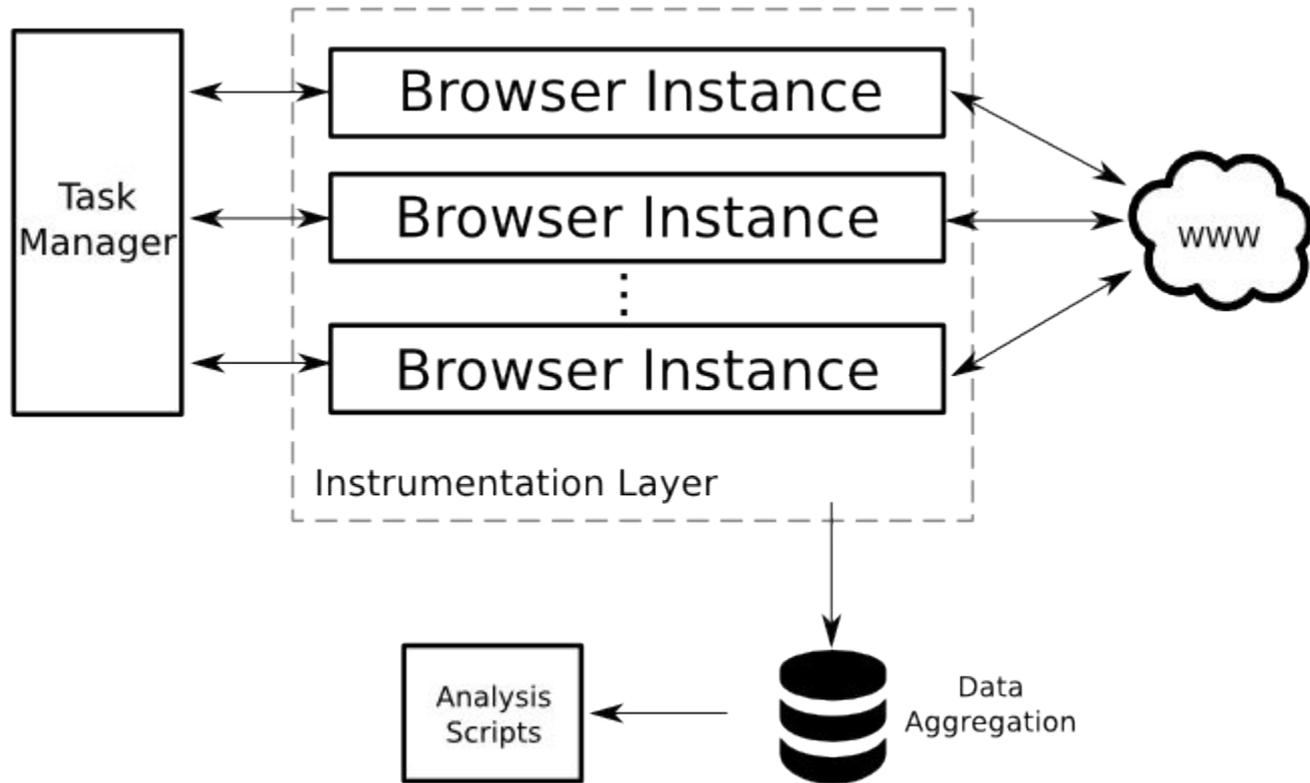


OpenWPM



OpenWPM





OpenWPM

- Supports browsing with persistent state
 - Browser keeps profile through crashes and freezes.
- Real Browser
 - Extensions
 - Privacy Features
 - WebRTC, Audio, Video, WebGL
- Stable

A Web Privacy Census

Monthly
1 Million Site Crawl

A Web Privacy Census

Monthly
1 Million Site Crawl

Collecting:

- Javascript Calls
- All javascript files
- HTTP Requests and Responses
- Storage (cookies, Flash, etc)

Targeted Crawls

Type	Use
Stateful	<ul style="list-style-type: none">● ID Cookies● Respawning● Cookie syncing
Stateless	<ul style="list-style-type: none">● Ghostery● AdBlock Plus● HTTPS Everywhere

A Web Privacy Census

1. Measure how effective tools are
2. Quickly deploy new measurements
3. Release data and analysis monthly

Detecting WebRTC Local IP Sniffing

2. I added code to JS Instrumentation for next crawl

```
// Access to webRTC
instrumentPrototype(window.mozRTCPeerConnection.prototype,
                    "mozRTCPeerConnection");
```

3. I wrote some analysis code

- Grab all urls that execute
 - `mozRTCPeerConnection.onicecandidate`
 - `mozRTCPeerConnection.createDataChannel`
 - `mozRTCPeerConnection.createOffer`
- Check JS Files to confirm

4. Results (October 2015)

- 121 first-party sites
 - 29 in the top 10k
- 24 unique scripts
- Only 1 of which is blocked by EasyList/EasyPrivacy

With regular measurement we can:

1. Inform the public
2. Build block lists
3. Change the incentives

2020

If you're going to track me, _____

2020

If you're going to track me, I'll know!

Help us make the web more private!

- **Contribute?**
 - github.com/citp/OpenWPM
- **Collaborate?**
 - webtap.princeton.edu

Image Assets from the Noun Project:

Microphone by Pavel N.; Megaphone by Piero Borgo; Smartphone by Aaron K. Kim; desktop computer and Databas by Creative Stall; link by Hash Basheer; Spider Bot by Siwat Vatatiyaporn; Browser by Dirtyworks; programmer by Hadi Davodpour