

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/> Google

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

Advertising **GUCCI** **Advertising** see what's inside

The New York Times Tuesday, February 1, 2011 Last Update: 10:16 PM ET

Search [select a foreign language](#) Follow Us [Subscribe to Home Delivery-Bay Area's Personalized News](#)

Switch to Global Edition

JOBS
REAL ESTATE
AUTOS
ALL CLASSIFIEDS

WORLD
U.S.
POLITICS
NEW YORK
BUSINESS
DEALBOOK
TECHNOLOGY
SPORTS
SCIENCE
HEALTH
OPINION
ARTS
Books
Movies
Music
Television
Theater
STYLE
Dining & Wine
Fashion & Style
Home & Garden
Weddings
Celebrations
TRAVEL

All Blogs
Cartoons
Classifieds
Corrections
Crossword / Games
Education
First Look
Learning Network
Multimedia
NYC Guide
Obituaries
Podcasts
Public Editor
Sunday Magazine
T Magazine
Video

Mubarak: Week New Term
Opposition Demands That He Leave Sooner

By [KAREEM FAHIM](#) and [ANTHONY SHADID](#) 54 minutes ago

Hundreds of thousands of Egyptians traveled like pilgrims to speak freely and to be heard.

Path to Change in Power Still Unclear
By [ANTHONY SHADID](#) 56 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)

Path to Change in Power Still Unclear
By [DAVID D. KIRKPATRICK](#) and [MARK LANDLER](#) 1 minute ago

The democracy movement is unfolding so rapidly in Egypt that Washington came close to being left behind.

Quiet Acts of Protest on a Noisy Day
By [KAREEM FAHIM](#) 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: Protests**

Video

OPINION
EGYPT
Kristof: The White House should support the protesters in Egypt.
Blog: Mubarak's Speech
Follow on Twitter
Brooks: In the democratic wave, a quest for dignity.
Cohen: From a culture of victimhood to one of self-empowerment.
Senator John Kerry: A future without Mubarak.

MORE IN OPINION
Editorial: Debt Limit
Bloggingheads: Abortions

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

MARKETS As of 10:14 PM EST

	Japan	Hong Kong	China
Nikkei	15,470.50	23,826.39	2,788.86
	+198.00	+343.44	Closed
	+1.01%	+1.48%	Hong Kong

Delayed at least 15 minutes

GET QUOTES [My Portfolio](#)
Stock, ETFs, Funds Go

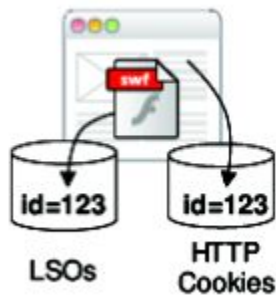
1.30 NEW FIELD SAVINGS ACCOUNT

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

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



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**

Panopticlick

How Unique — and Trackable — Is Your Browser?

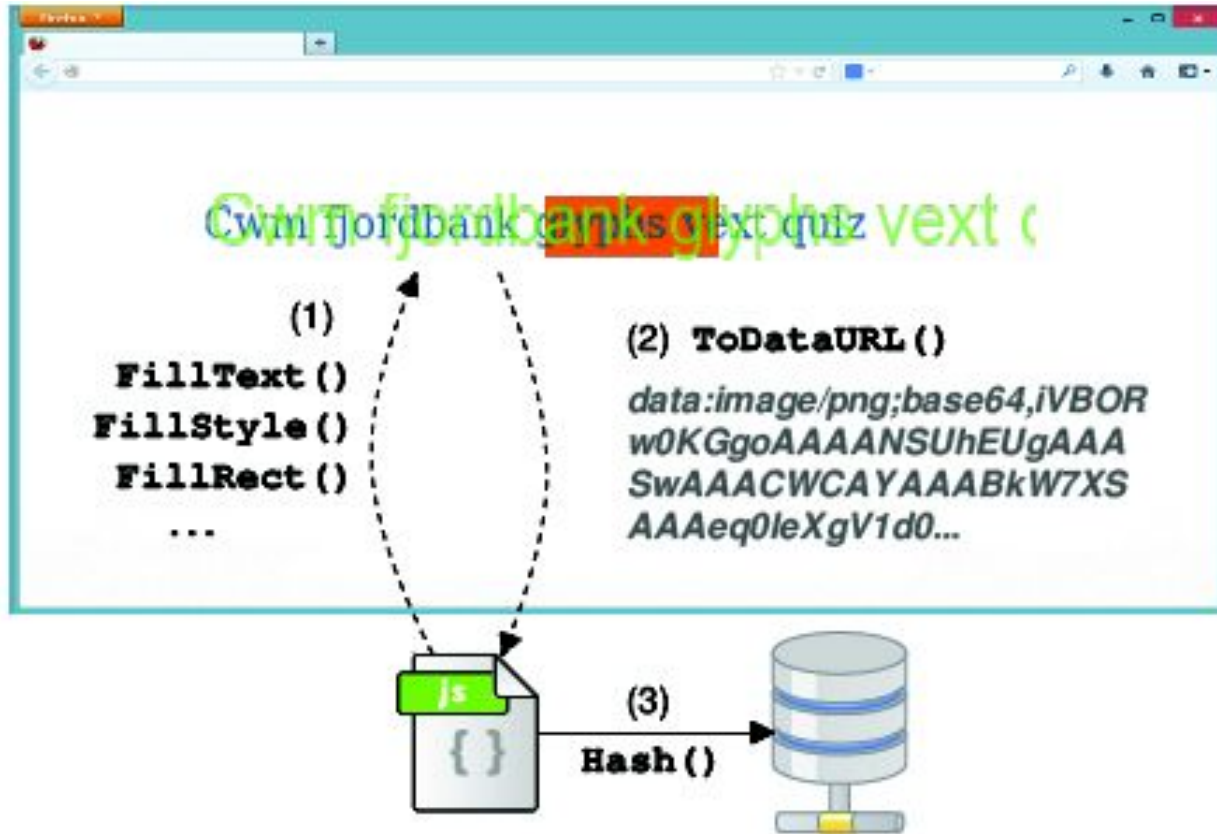
Is your browser configuration rare or unique? If so, web sites may be able to track you, *even if you limit or disable cookies.*

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.

TEST
ME

Canvas Fingerprinting



2009

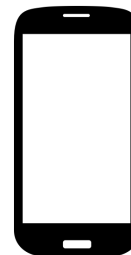
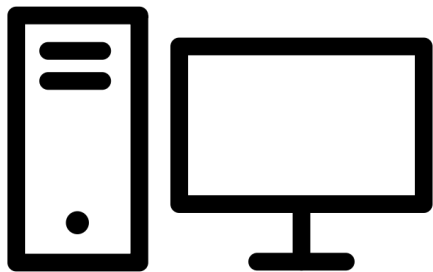
*If you're going to track me, please **use cookies***

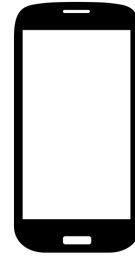
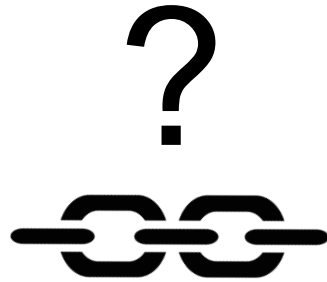
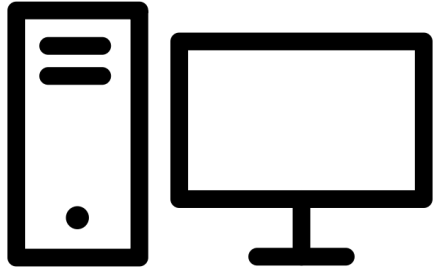
2009

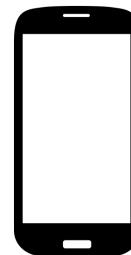
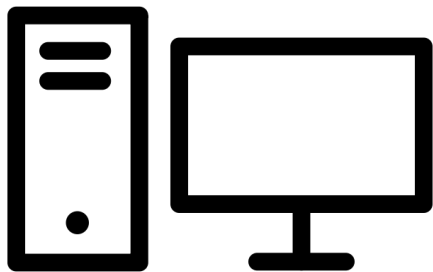
~~*If you're going to track me, please **use cookies***~~

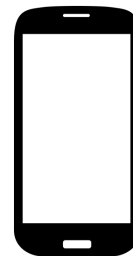
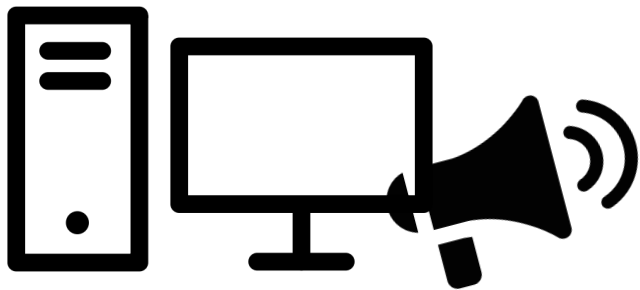
2010

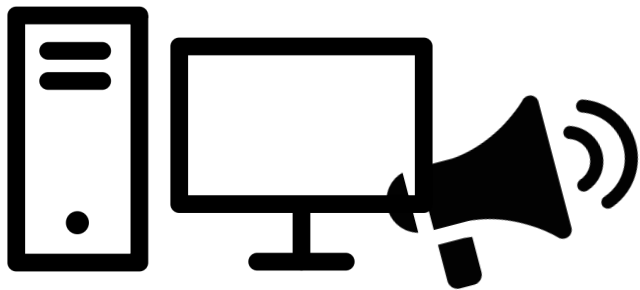
*If you're going to track me, please **use browser storage***













FEDERAL TRADE COMMISSION

PROTECTING AMERICA'S CONSUMERS

[Contact](#)[ABOUT THE FTC](#)[NEWS & EVENTS](#)[ENFORCEMENT](#)[POLICY](#)[TIPS & ADV](#)

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

Cross-Device Tracking



NOV 16, 2015

CONSTITUTION CENTER

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

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

Re

Ma

FT

Tr

No

FT

for

Wc

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***

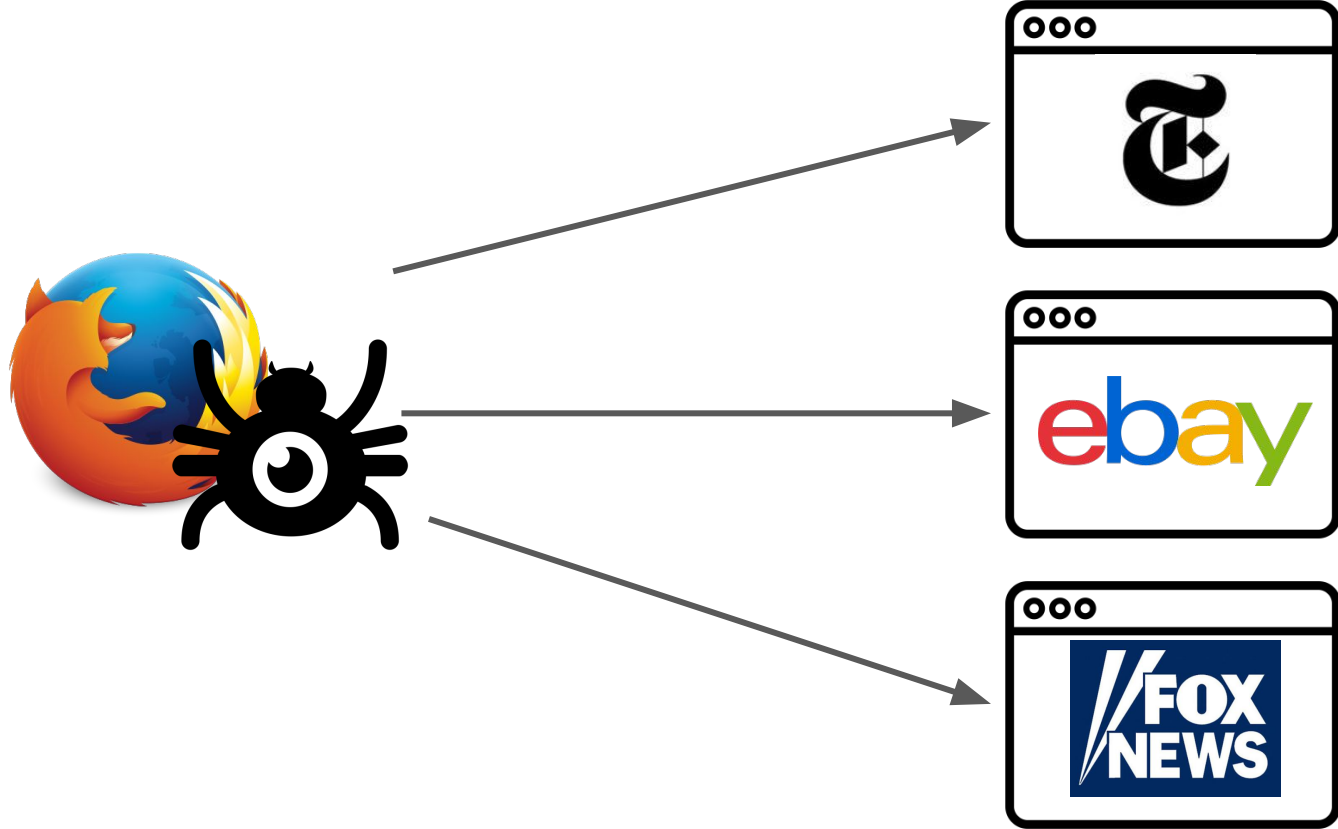
2015

~~*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	Automation ^{FF}	Infrastructure Instrumentation	Variable	Scale
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}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/CapsperJS (Headless webkit)

- Instrumentation

- 5 used FourthParty
- 9 used a Proxy

Many Studies, Many Platforms

- Automation:
 - 7 used Selenium (Full browser)
 - 4 used PhantomJS/CapsperJS (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

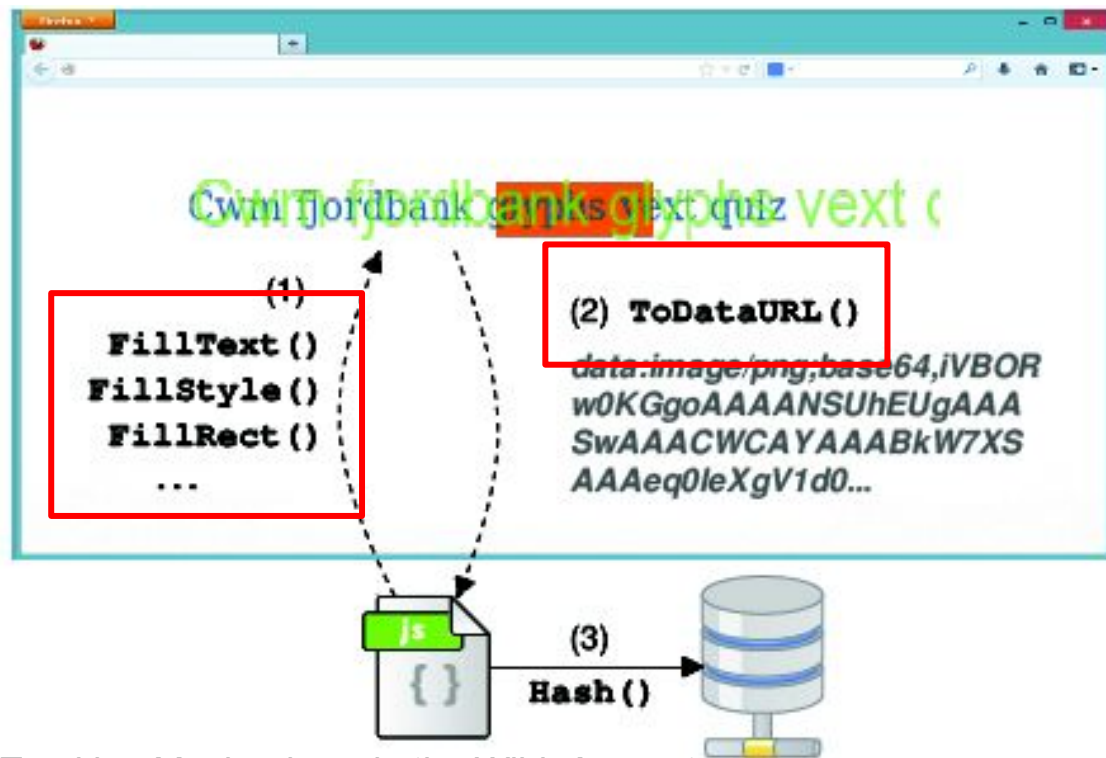
- Acar, et.al (2014)
- 5% of Top 100k

- [illegible]

The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. Acar, et.al.

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

Canva


Over

The W

esupply.com


Digital Delivery in Seconds!

Game Cards | Gift Cards | PrePaid Cards




Blocked 2 potential HTML canvas fingerprinting attempts on this page

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



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



[Donate](#)

15

100k

Overcoming these hurdles:

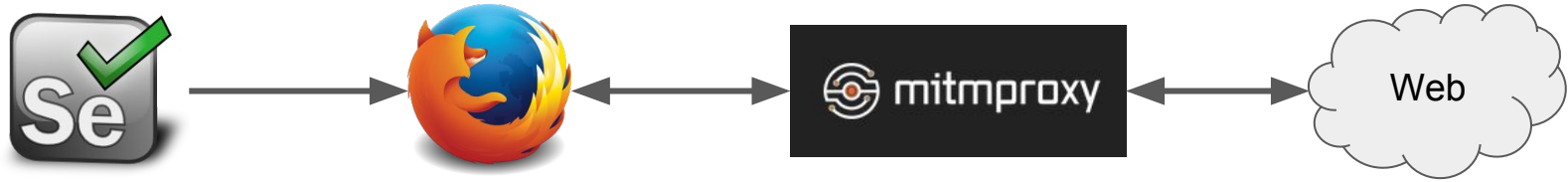
1. A Common Platform

2. A Web Privacy Census

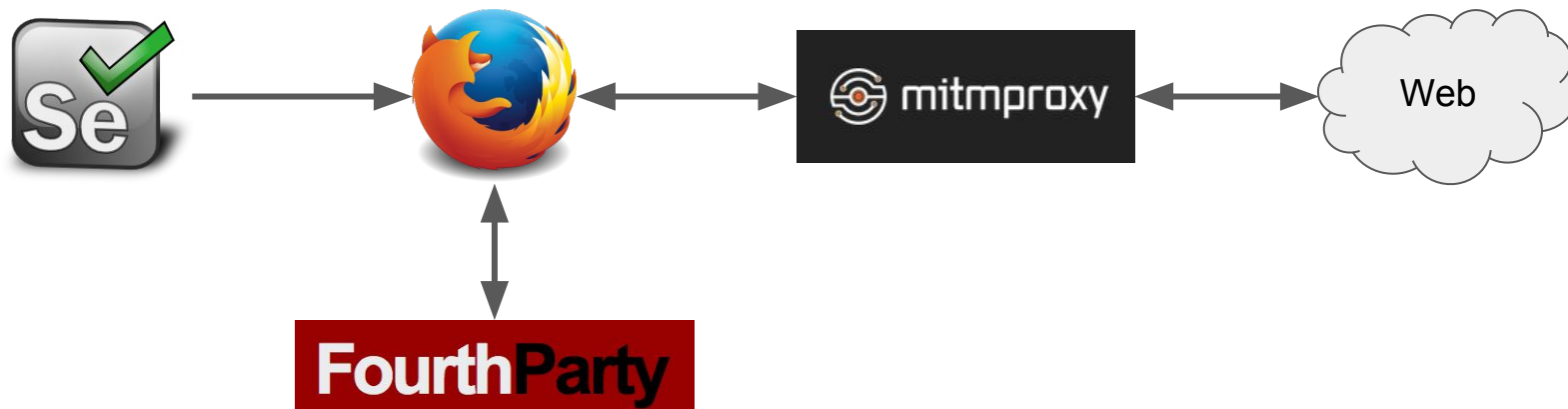
OpenWPM



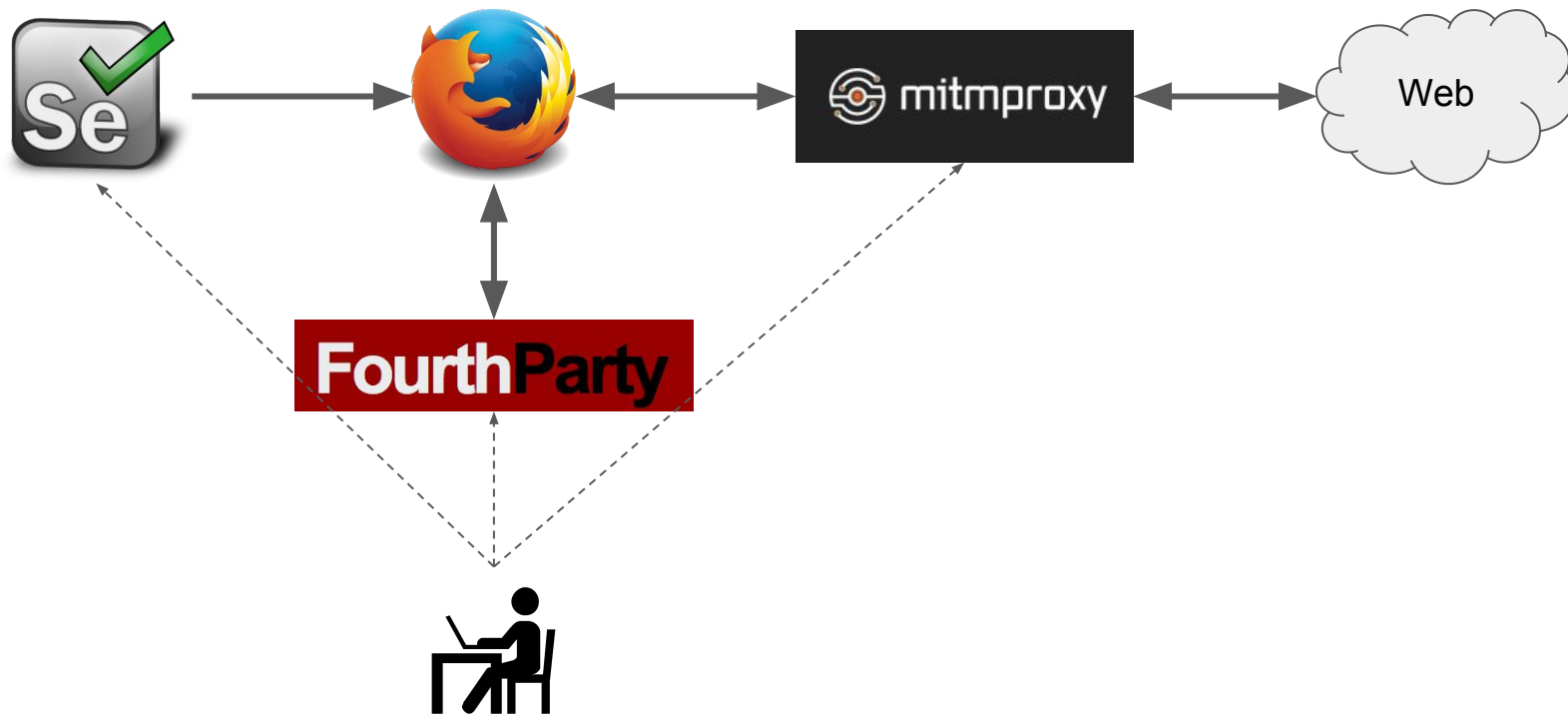
OpenWPM



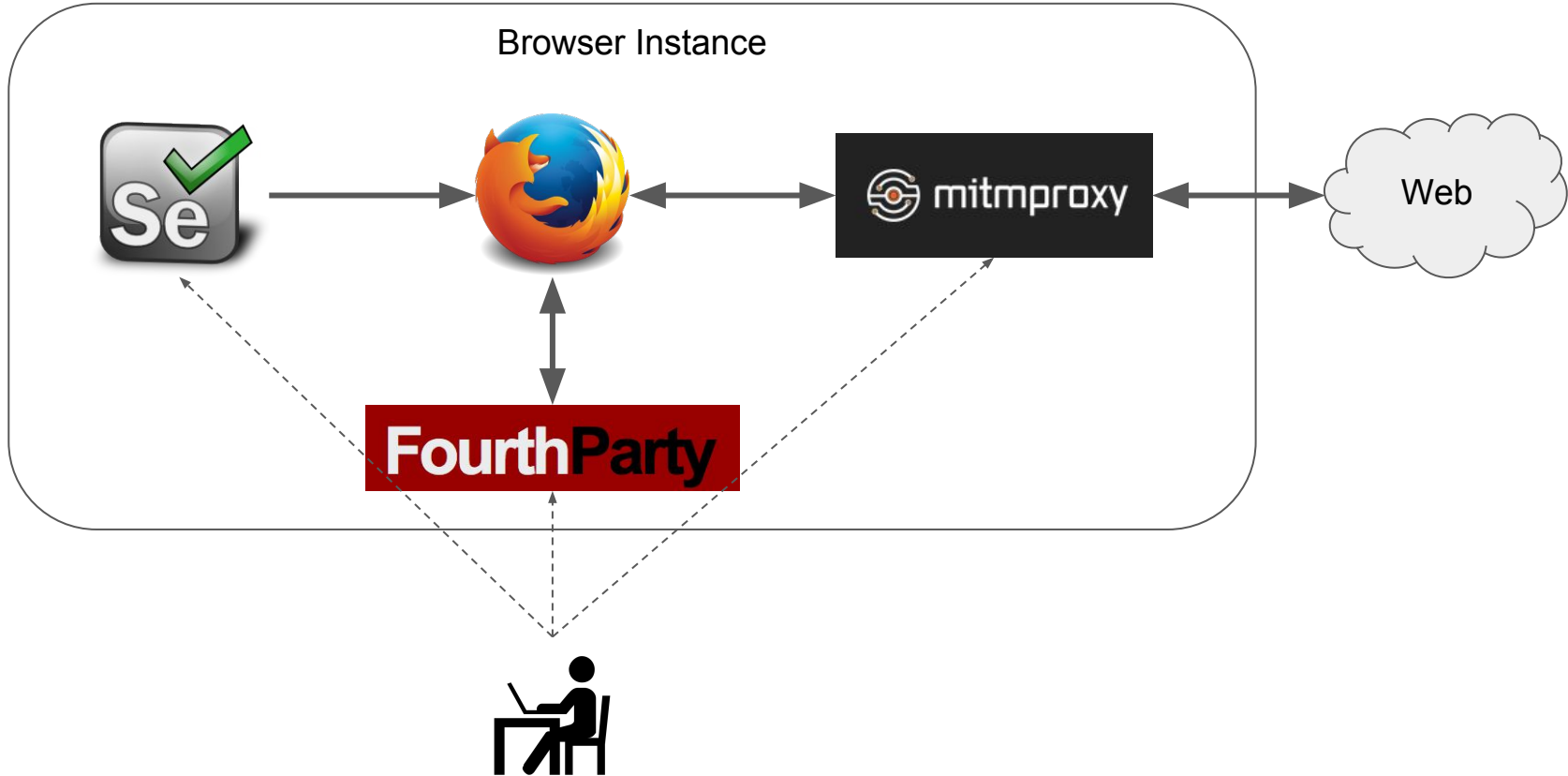
OpenWPM

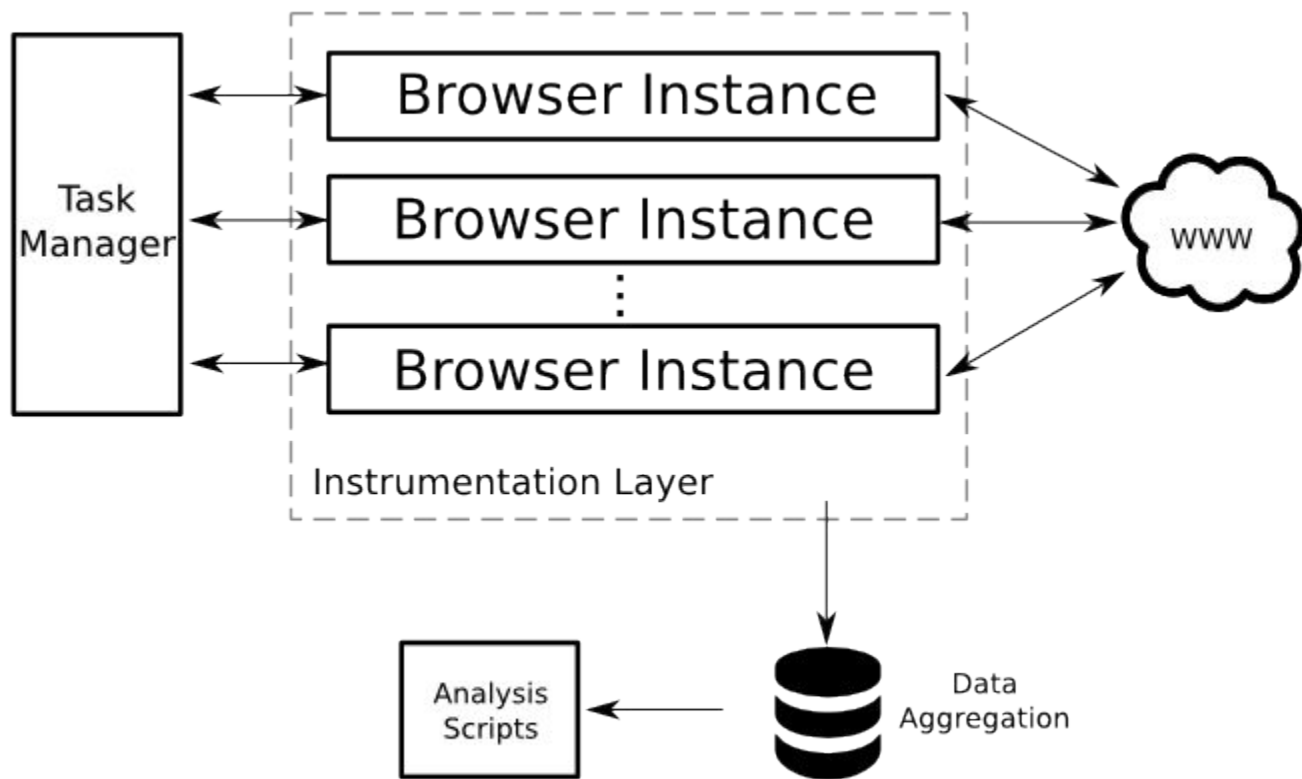


OpenWPM



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	<ul style="list-style-type: none">● 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

1. I saw a tweet that nytimes.com is IP sniffing



Mike O'Neill

@incloud



Follow

WebRTC being used now by embedded 3rd party on nytimes.com to report visitors' local IP addresses.

```
Debugger - http://www.nytimes.com/?WT.z_jog=1&hf
Vector Console Debugger Style Editor Performance Network DOM
8.nytimes.com
kats.on.nytime...
red2.google...
m3
m3.googleads...
4.63/j
ads.g.doublecl...
parvix.com
143871988967000
[...]
```

Time	IP Address	IP Address	IP Address	Operation
148	0.0062285	192.168.1.203	192.168.1.254	DNS Query Operation,
143		192.168.1.119	239.255.255.250	SSDP Request, Method:
571	0.00000005	192.168.1.203	192.168.1.254	DNS Request, Query I
573	0.0148223	192.168.1.203	192.168.1.254	DNS Query Operation,
712		192.168.1.254	192.168.1.203	DNS Response, #Code:
210		192.168.1.107	239.255.255.250	SSDP Request, Method:
359		192.168.1.107	239.255.255.250	SSDP Request, Method:
406		192.168.1.107	239.255.255.250	SSDP Request, Method:
529	0.0268478	192.168.1.203	192.168.1.254	DNS Query Operation,
808	0.00000005	192.168.1.203	192.168.1.254	DNS Request, Query I
906	0.0053289	192.168.1.203	192.168.1.254	DNS Query Operation,
210		192.168.1.254	192.168.1.203	DNS Response, #Code:
336	0.0282345	192.168.1.203	192.168.1.254	DNS Query Operation,
420	0.0137929	192.168.1.203	192.168.1.254	DNS Query Operation,
480	0.0118086	192.168.1.203	192.168.1.254	DNS Query Operation,

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