Online Tracking

A 1-million-site Measurement and Analysis

Steven Englehardt
@s_englehardt

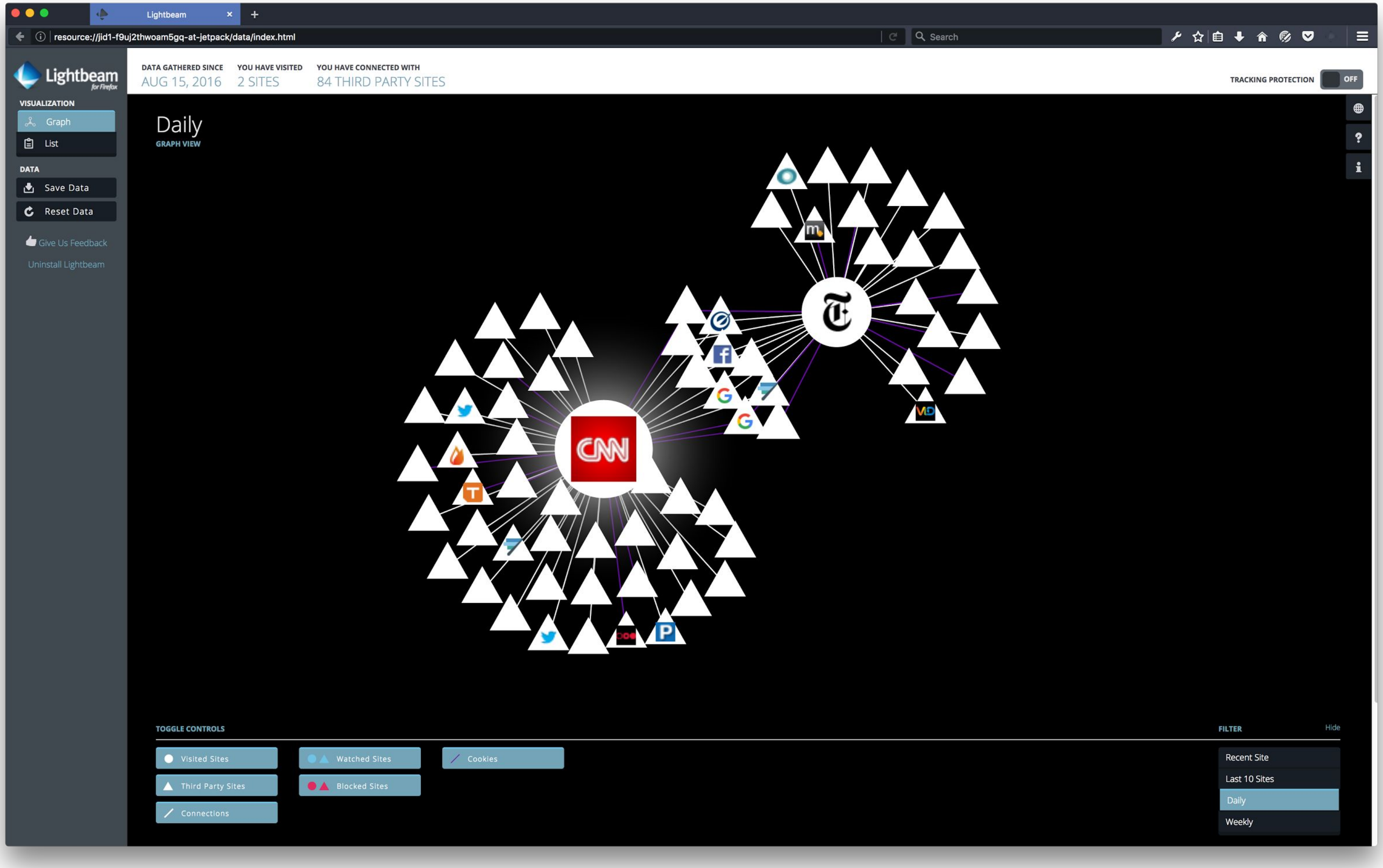
Arvind Narayanan
@random_walker

PRINCETON
UNIVERSITY

CENTER FOR
INFORMATION
TECHNOLOGY
POLICY
PRINCETON UNIVERSITY

This research was supported by NSF award CNS 1526353, a grant from the Data Transparency Lab, and an Amazon AWS Credits Research Grant.

Visiting 2 websites results in 84 third parties contacted

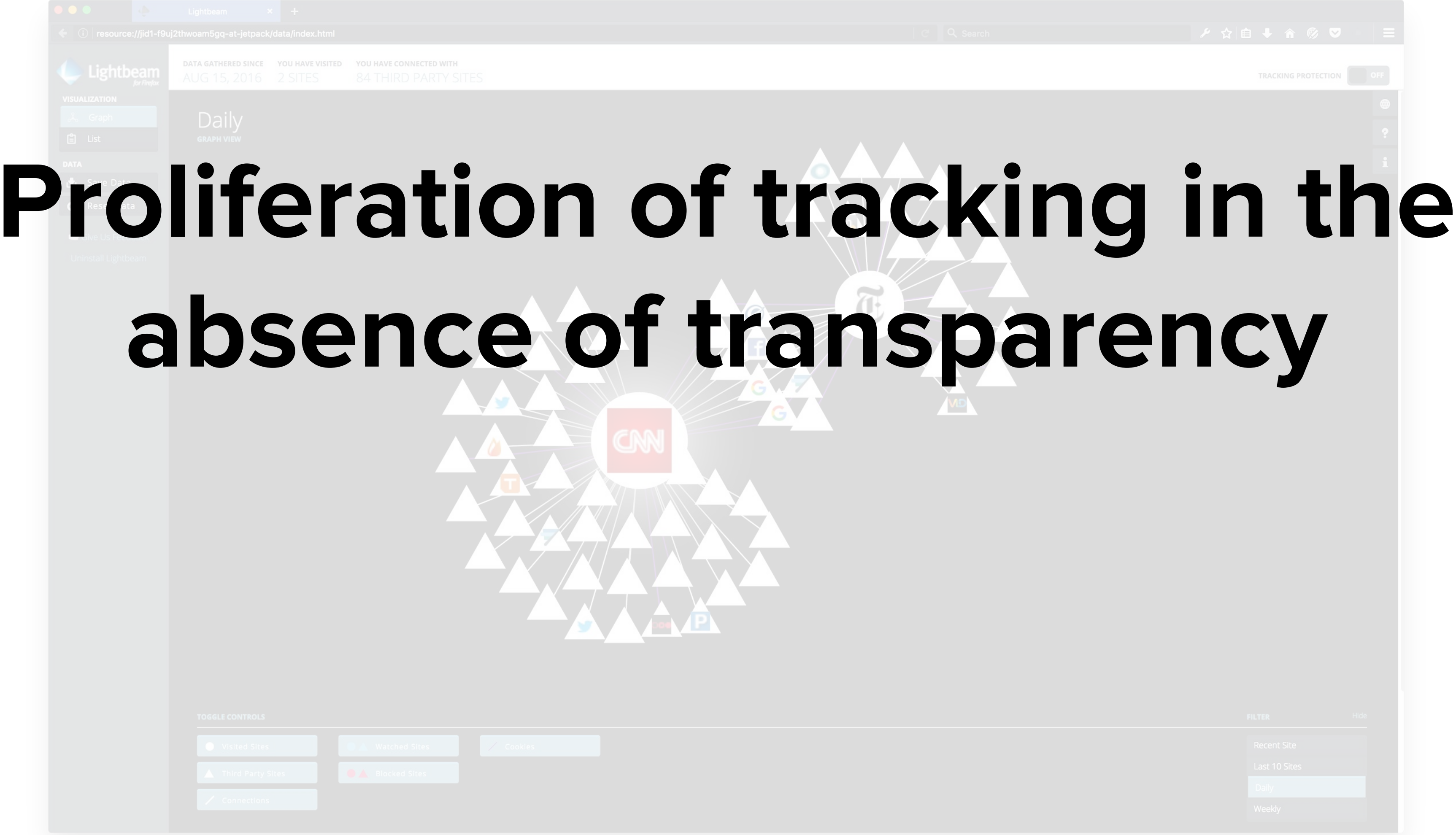


DATA GATHERED SINCE AUG 15, 2016

YOU HAVE VISITED 2 SITES

YOU HAVE CONNECTED WITH 84 THIRD PARTY SITES

Visiting 2 websites results in 84 third parties contacted



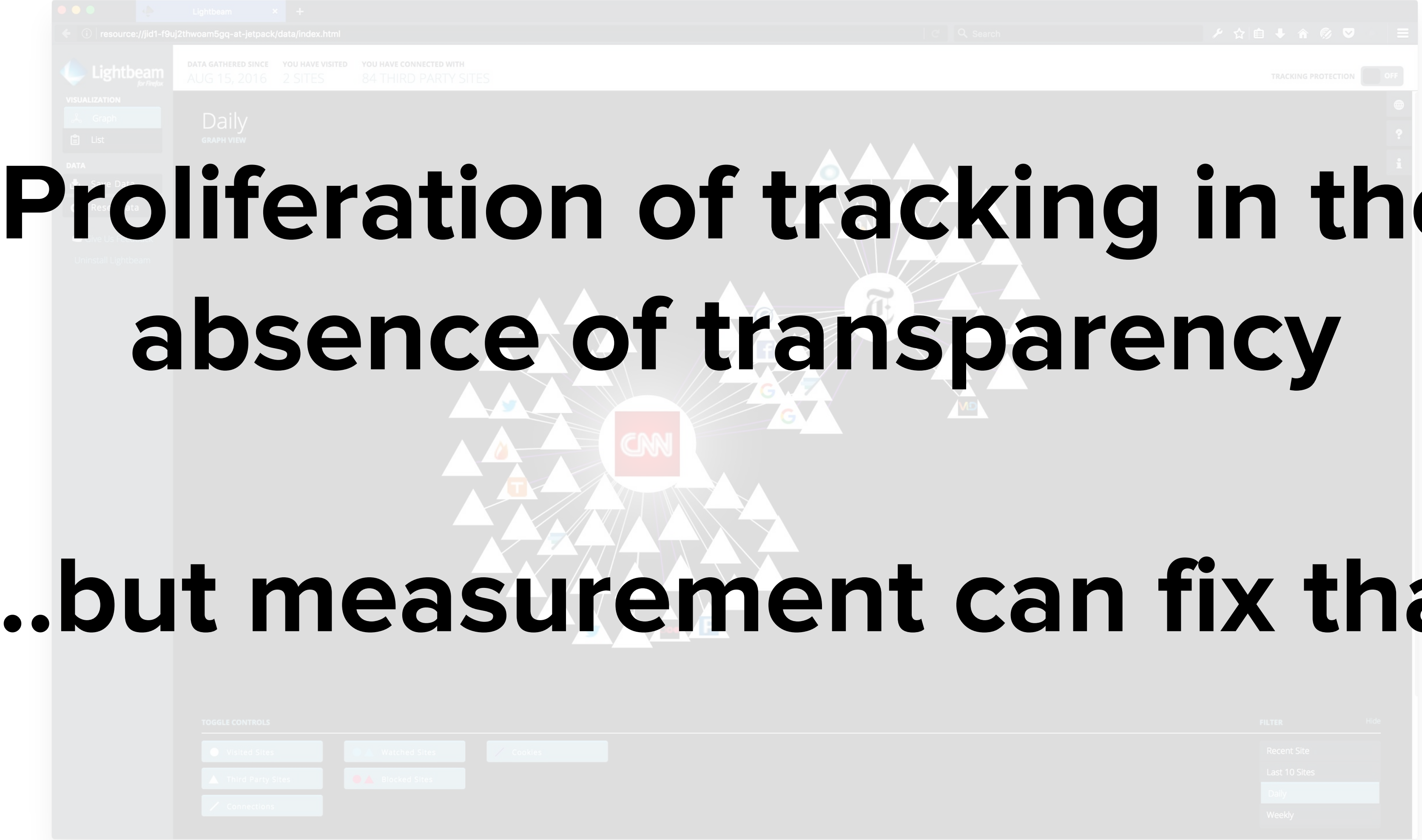
DATA GATHERED SINCE AUG 15, 2016

YOU HAVE VISITED 2 SITES

YOU HAVE CONNECTED WITH 84 THIRD PARTY SITES

Proliferation of tracking in the absence of transparency

Visiting 2 websites results in 84 third parties contacted



resource://jid1-f9uj2thwoam5gq-at-jetpack/data/index.html

DATA GATHERED SINCE AUG 15, 2016

YOU HAVE VISITED 2 SITES

YOU HAVE CONNECTED WITH 84 THIRD PARTY SITES

Proliferation of tracking in the absence of transparency

...but measurement can fix that

Measurement forces companies to fix problems

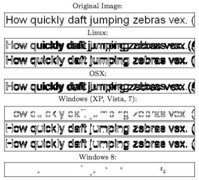


Figure 7: Difference maps for a group on `text_arial`

Canvas
Fingerprinting
Introduced

Mowery and Shacham (W2SP 2012)

May 2012

Measurement forces companies to fix problems

Figure 7: Difference maps for a group on text_arial

Canvas Fingerprinting Introduced

Mowery and Shacham (W2SP 2012)

AddThis

LIGATUS

May 2012

Canvas fingerprinting adopted over 2 years

Measurement forces companies to fix problems

Original Image:
How quickly daft jumping zebras vex. (
Linux:
OSX:
Windows (XP, Vista, 7):
Windows 8:
Figure 7: Difference maps for a group on text_arial

Canvas Fingerprinting Introduced

Mowery and Shacham (W2SP 2012)

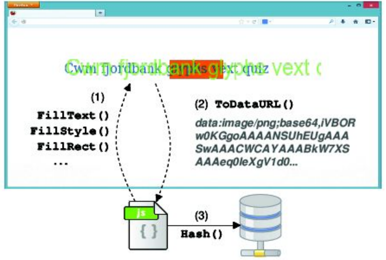
May 2012

Canvas fingerprinting adopted over 2 years

News Coverage

July 21st 2014

Measurement Results Released



The Web Never Forgets: Persistent Tracking Mechanisms in the Wild (CCS 2014)

Measurement forces companies to fix problems

Figure 7: Difference maps for a group on `text_arial`

Canvas
Fingerprinting
Introduced

Mowery and Shacham (W2SP 2012)

AddThis

LIGATUS

News
Coverage

ProPublica

BBC

SPIEGEL ONLINE

July 21st
2014

AddThis

LIGATUS

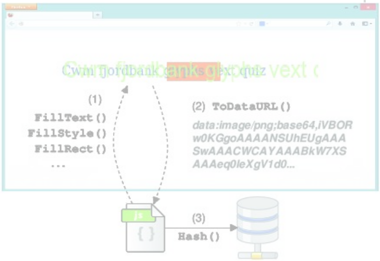
Largest
Fingerprinters
Stopped

May 2012

Canvas fingerprinting adopted
over 2 years

July 23rd
2014

Measurement
Results
Released



The Web Never Forgets: Persistent Tracking Mechanisms in the Wild (CCS 2014)

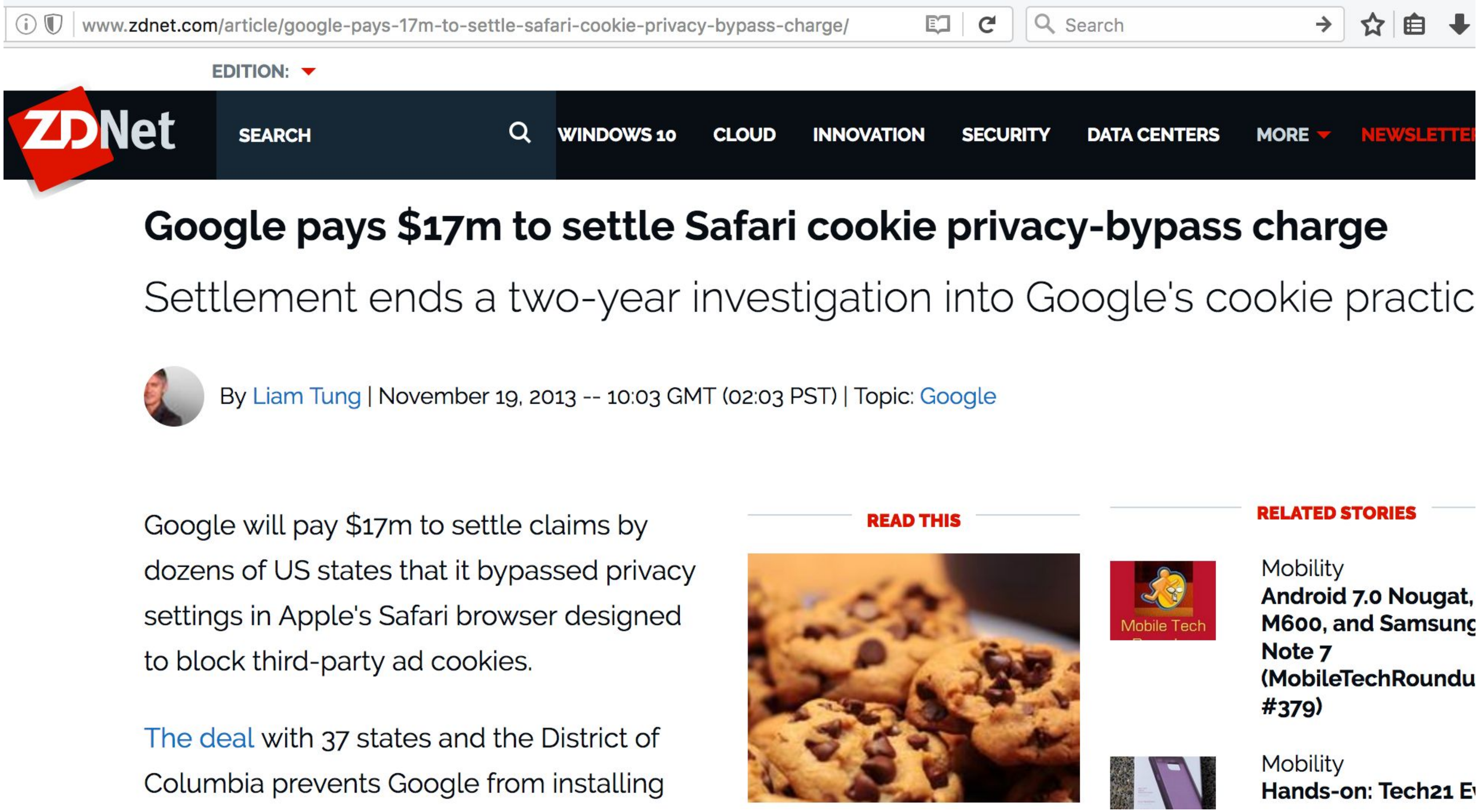
Measurement is effective because most actors are not malicious

1. Bulk of trackers respond to pressure from publishers, users, and regulators

2. Few instances of trying to avoid detection

3. High risk for malicious actions

Google settlement for subverting cookie blocking



Google pays $17m to settle Safari cookie privacy-bypass charge

Settlement ends a two-year investigation into Google's cookie practic

By Liam Tung | November 19, 2013 -- 10:03 GMT (02:03 PST) | Topic: Google

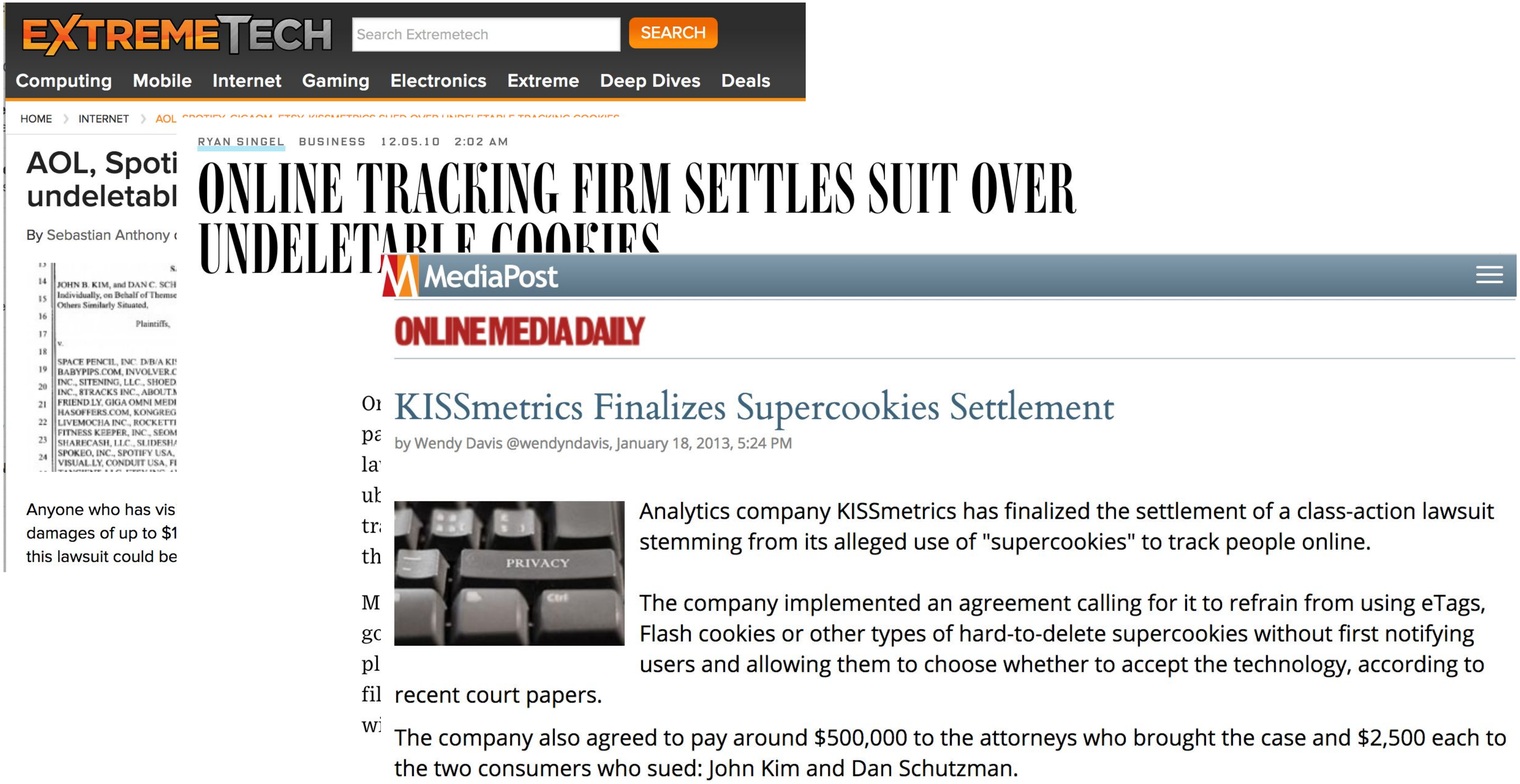
Google will pay $17m to settle claims by dozens of US states that it bypassed privacy settings in Apple's Safari browser designed to block third-party ad cookies.

The deal with 37 states and the District of Columbia prevents Google from installing

Mobility
Android 7.0 Nougat, M600, and Samsung Note 7 (MobileTechRoundu #379)

Mobility
Hands-on: Tech21 E

Multiple settlements for subverting cookie clearing



AOL, Spoti
undeletabl

By Sebastian Anthony o

S.

JOHN B. KIM, and DAN C. SCH
Individually, on Behalf of Themse
Others Similarly Situated,

Plaintiffs,

v.

SPACE PENCIL, INC. D/B/A KI
BABYPIPS.COM, INVOLVER.C
INC., SITENING, LLC., SHOED
INC., 8TRACKS INC., ABOUT.M
FRIEND.LY, GIGA OMNI MEDI
HASOFFERS.COM, KONGREG
LIVEMOCHA INC., ROCKETTI
FITNESS KEEPER, INC., SEOM
SHARECASH, LLC., SLIDESHA
SPOKEO, INC., SPOTIFY USA,
VISUAL.LY, CONDUIT USA, FI

Anyone who has vis
damages of up to $1
this lawsuit could be

RYAN SINGEL BUSINESS 12.05.10 2:02 AM

ONLINE TRACKING FIRM SETTLES SUIT OVER UNDELETABLE COOKIES

On
pa
la
ub
tra
th

M
go
pl
fil
wi

KISSmetrics Finalizes Supercookies Settlement

by Wendy Davis @wendyndavis, January 18, 2013, 5:24 PM

Analytics company KISSmetrics has finalized the settlement of a class-action lawsuit stemming from its alleged use of "supercookies" to track people online.

The company implemented an agreement calling for it to refrain from using eTags, Flash cookies or other types of hard-to-delete supercookies without first notifying users and allowing them to choose whether to accept the technology, according to recent court papers.

The company also agreed to pay around $500,000 to the attorneys who brought the case and $2,500 each to the two consumers who sued: John Kim and Dan Schutzman.

Flash Cookies and Privacy (2009) Soltani, et al.
Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning (2011) Ayenson, et al.

Automated, large-scale measurement returns control to users and publishers

1. **Our measurement platform**

2. Insights from our 1-million-site measurement

3. Next steps

	Targets	Infrastructure				Variable					
Paper		Automation[a]	Instrumentation	Crowd-sourced	Distributed	Location	User-agent	Demographics	Interests	Privacy Tools	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

[a] 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

A need for a common platform

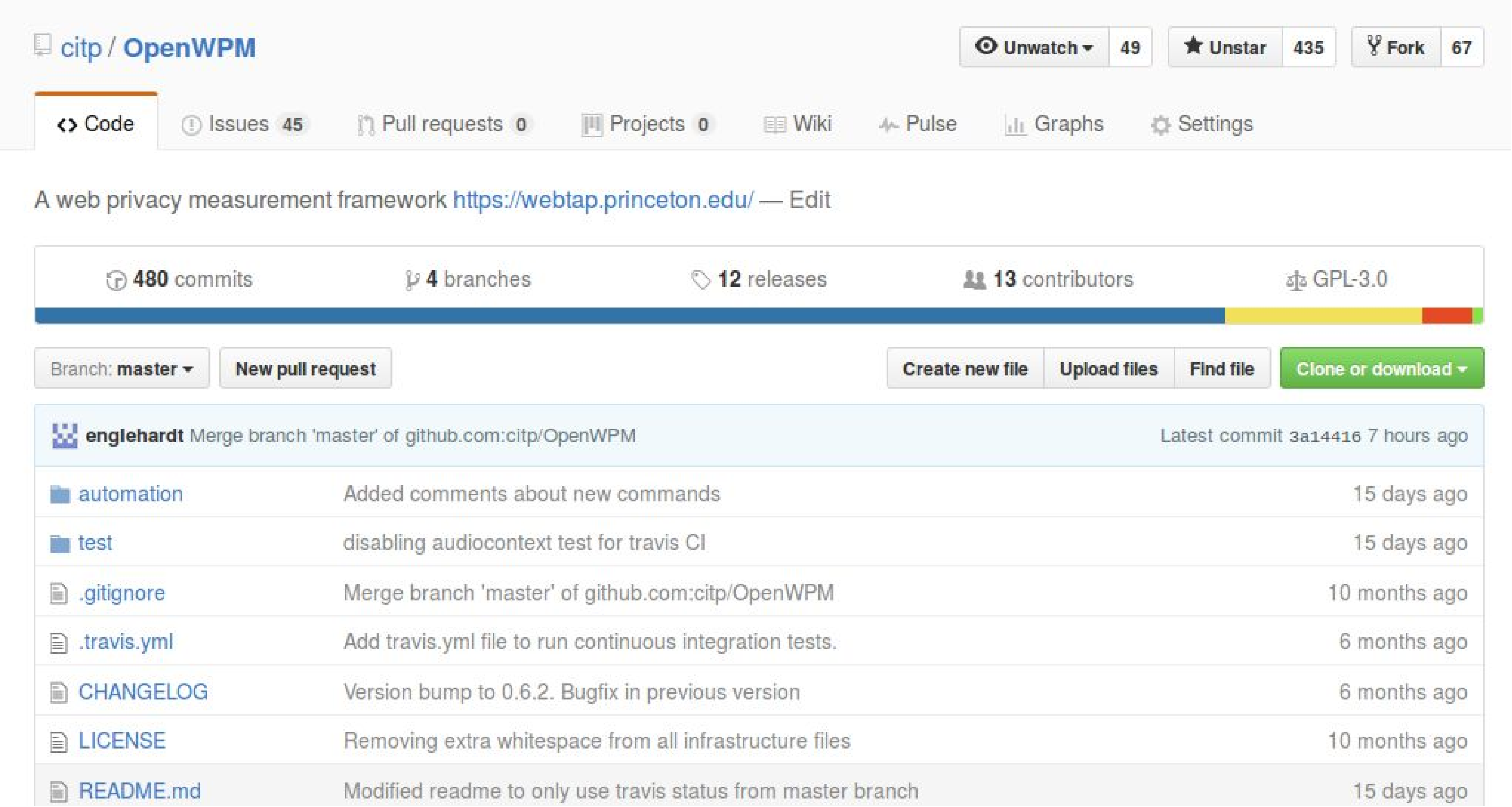
- Re-engineering of similar measurement tools
- Methodological differences between platforms
  - PhantomJS vs Firefox vs Chrome
- High cost to reproduce or re-measure
  - Studies are only run once
- Can build upon other open measurement tools

FourthParty -- *Third-party web tracking: Policy and technology* -- Mayer et al. 2012

FPDetective -- *FPDetective: dusting the web for fingerprinters* -- Acar et al. 2013

Chameleon -- *https://github.com/ghostwords/chameleon*

Our Web Privacy Measurement (WPM) Platform



A web privacy measurement framework https://webtap.princeton.edu/ — Edit

480 commits · 4 branches · 12 releases · 13 contributors · GPL-3.0

englehardt Merge branch 'master' of github.com:citp/OpenWPM · Latest commit `3a14416` 7 hours ago

automation	Added comments about new commands	15 days ago
test	disabling audiocontext test for travis CI	15 days ago
.gitignore	Merge branch 'master' of github.com:citp/OpenWPM	10 months ago
.travis.yml	Add travis.yml file to run continuous integration tests.	6 months ago
CHANGELOG	Version bump to 0.6.2. Bugfix in previous version	6 months ago
LICENSE	Removing extra whitespace from all infrastructure files	10 months ago
README.md	Modified readme to only use travis status from master branch	15 days ago

https://github.com/citp/OpenWPM

Study using OpenWPM	**Conf.**	**Year**
The Web Never Forgets: Persistent Tracking Mechanisms in the Wild	**CCS**	**2014**
Cognitive disconnect:Understanding Facebook Connect login permissions	OSN	2014
Cookies that give you away: The surveillance implications of web tracking	**WWW**	**2015**
Upgrading HTTPS in midair: HSTS and key pinning in practice	NDSS	2015
Web Privacy Census	Tech Science	2015
Variations in Tracking in Relation to Geographic Location	W2SP	2015
No Honor Among Thieves: A Large-Scale Analysis of Malicious Web Shells	WWW	2016
Online Tracking: A 1-million-site Measurement and Analysis	**CCS**	**2016**
Dial One for Scam: Analyzing and Detecting Technical Support Scams	[Working Paper]	2016

Study using OpenWPM	Conf.	Year
The Web Never Forgets: Persistent Tracking Mechanisms in the Wild	**CCS**	**2014**
Cognitive disconnect:Understanding Facebook Connect login permissions	OSN	2014
Cookies that give you away: The surveillance implications of web tracking	**WWW**	**2015**
Upgrading HTTPS in midair: HSTS and key pinning in practice	NDSS	2015
Web Privacy Census	Tech Science	2015
Variations in Tracking in Relation to Geographic Location	W2SP	2015
No Honor Among Thieves: A Large-Scale Analysis of Malicious Web Shells	WWW	2016
Online Tracking: A 1-million-site Measurement and Analysis	**CCS**	**2016**
Dial One for Scam: Analyzing and Detecting Technical Support Scams	[Working Paper]	2016

1. Our measurement platform

2. **Insights from our 1-million-site measurement**

3. Next steps

Insights from our 1-million-site measurement

1. There is a long but thin talk of tracker presence on the top sites.
2. Develop a metric to rank tracker popularity.
3. Show that third-parties (and trackers) impede HTTPS adoption and cause mixed content warnings
4. Evaluate differences in tracking across categories (e.g. news sites >>> adult)
5. Examine how common cookie syncing is
6. Measure the use of the HTML Canvas for fingerprinting
7. Measure several HTML5 fingerprinting techniques
8. Examine how well tracking protection detects trackers

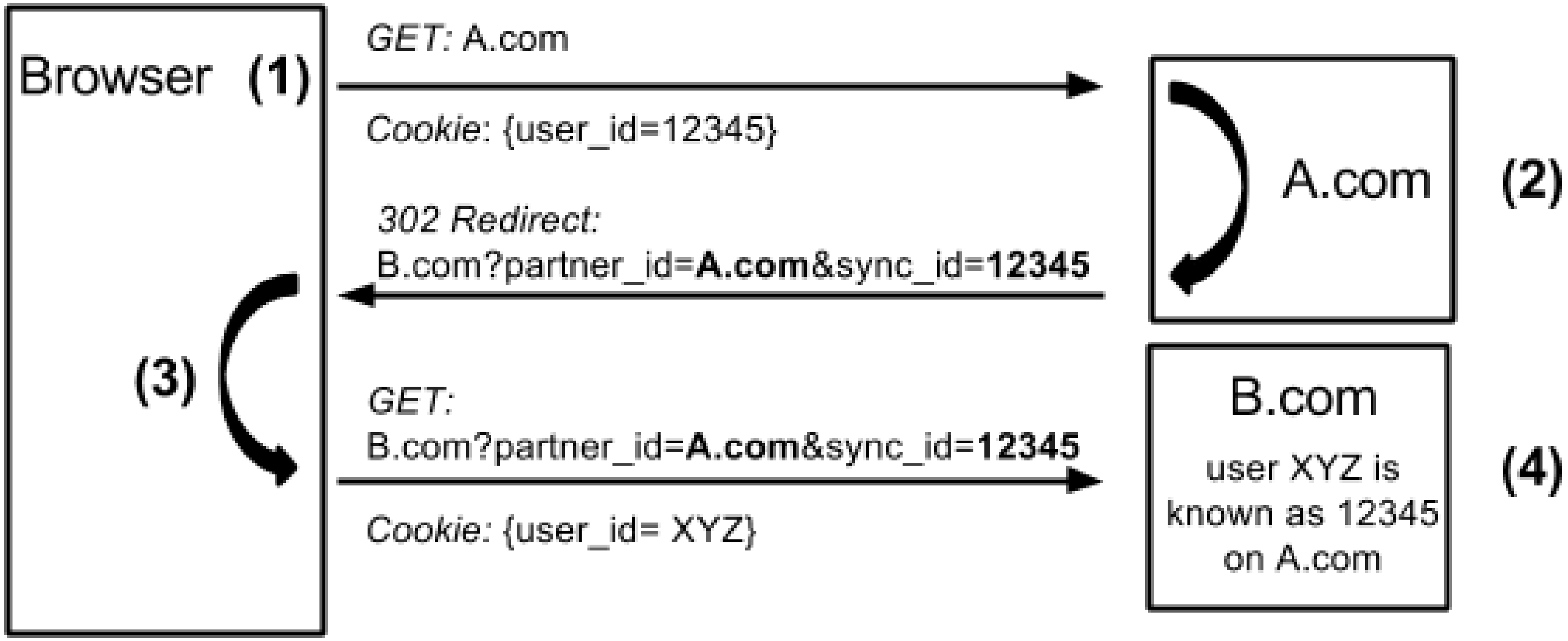
Full Paper: senglehardt.com/papers/ccs16_online_tracking.pdf

Insights from our 1-million-site measurement

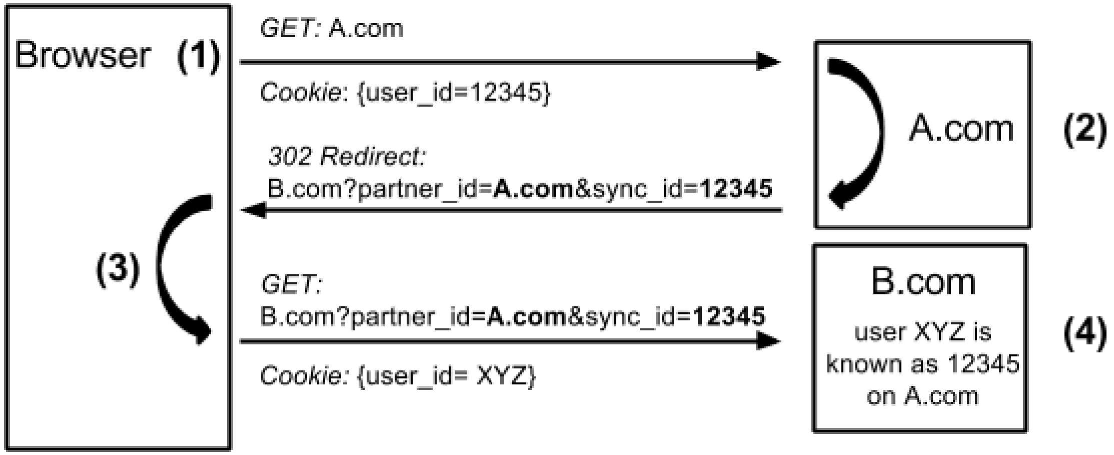
1. There is a long but thin talk of tracker presence on the top sites.
2. Develop a metric to rank tracker popularity.
3. Show that third-parties (and trackers) impede HTTPS adoption and cause mixed content warnings
4. Evaluate differences in tracking across categories (e.g. news sites >>> adult)
5. Examine how common cookie syncing is
6. Measure the use of the HTML Canvas for fingerprinting
7. Measure several HTML5 fingerprinting techniques
8. Examine how well tracking protection detects trackers

Full Paper: senglehardt.com/papers/ccs16_online_tracking.pdf

Almost all top third parties cookie sync



Almost all top third parties cookie sync

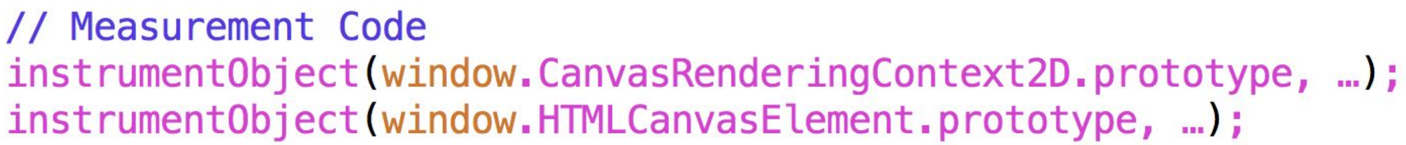

Browser (1)
GET: A.com
Cookie: {user_id=12345}
A.com (2)
302 Redirect:
B.com?partner_id=A.com&sync_id=12345
(3)
GET:
B.com?partner_id=A.com&sync_id=12345
Cookie: {user_id= XYZ}
B.com
user XYZ is known as 12345 on A.com
(4)

45 of top 50 third parties sync cookies (85% chance any two share an ID)

85 of the top 100 (66% chance any two share an ID)

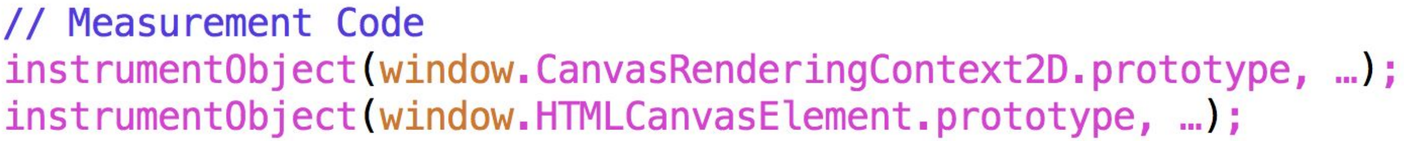
Several HTML5 Features Used for Fingerprinting

Detecting Fingerprinting

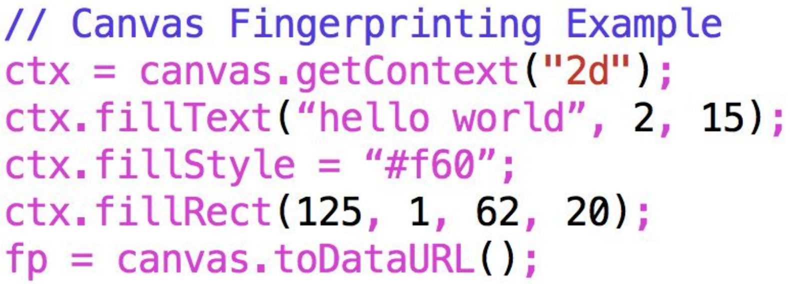


```
// Measurement Code
instrumentObject(window.CanvasRenderingContext2D.prototype, …);
instrumentObject(window.HTMLCanvasElement.prototype, …);
```

Detecting Fingerprinting

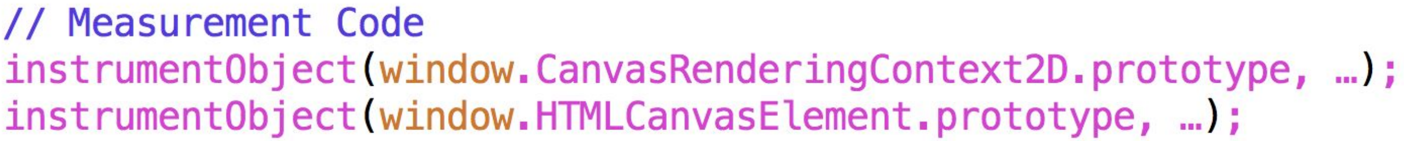


```
// Measurement Code
instrumentObject(window.CanvasRenderingContext2D.prototype, …);
instrumentObject(window.HTMLCanvasElement.prototype, …);
```

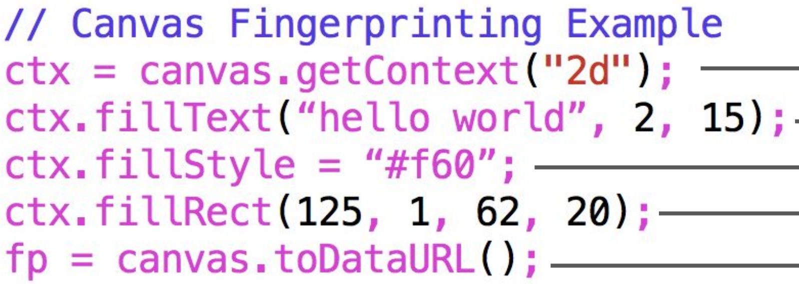


```
// Canvas Fingerprinting Example
ctx = canvas.getContext("2d");
ctx.fillText(“hello world”, 2, 15);
ctx.fillStyle = “#f60”;
ctx.fillRect(125, 1, 62, 20);
fp = canvas.toDataURL();
```

Detecting Fingerprinting



```
// Measurement Code
instrumentObject(window.CanvasRenderingContext2D.prototype, …);
instrumentObject(window.HTMLCanvasElement.prototype, …);
```



```
// Canvas Fingerprinting Example
ctx = canvas.getContext("2d");
ctx.fillText(“hello world”, 2, 15);
ctx.fillStyle = “#f60”;
ctx.fillRect(125, 1, 62, 20);
fp = canvas.toDataURL();
```

Measurement Logs

(SCRIPT_URL, “getContext”, “2d”)

(SCRIPT_URL, “fillText”, “hello world”, 2, 15)

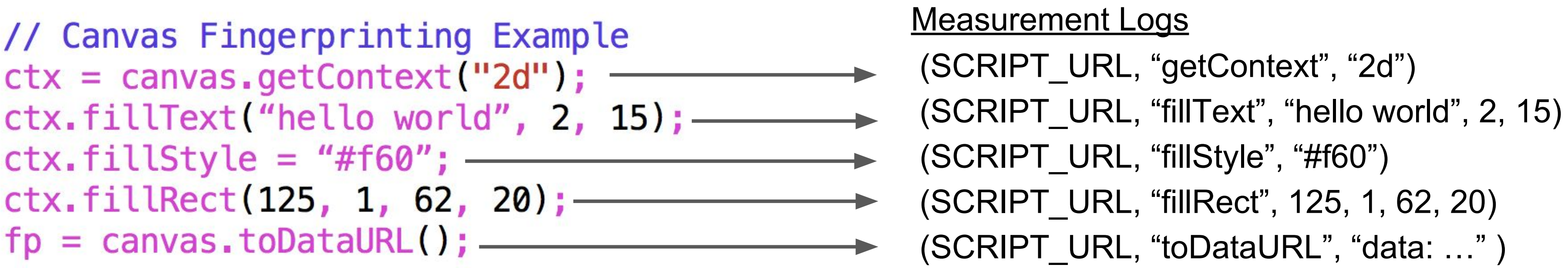
(SCRIPT_URL, “fillStyle”, “#f60”)

(SCRIPT_URL, “fillRect”, 125, 1, 62, 20)

(SCRIPT_URL, “toDataURL”, “data: …”)

Detecting Fingerprinting

```
// Measurement Code
instrumentObject(window.CanvasRenderingContext2D.prototype, …);
instrumentObject(window.HTMLCanvasElement.prototype, …);
```



```
// Canvas Fingerprinting Example
ctx = canvas.getContext("2d");
ctx.fillText(“hello world”, 2, 15);
ctx.fillStyle = “#f60”;
ctx.fillRect(125, 1, 62, 20);
fp = canvas.toDataURL();
```

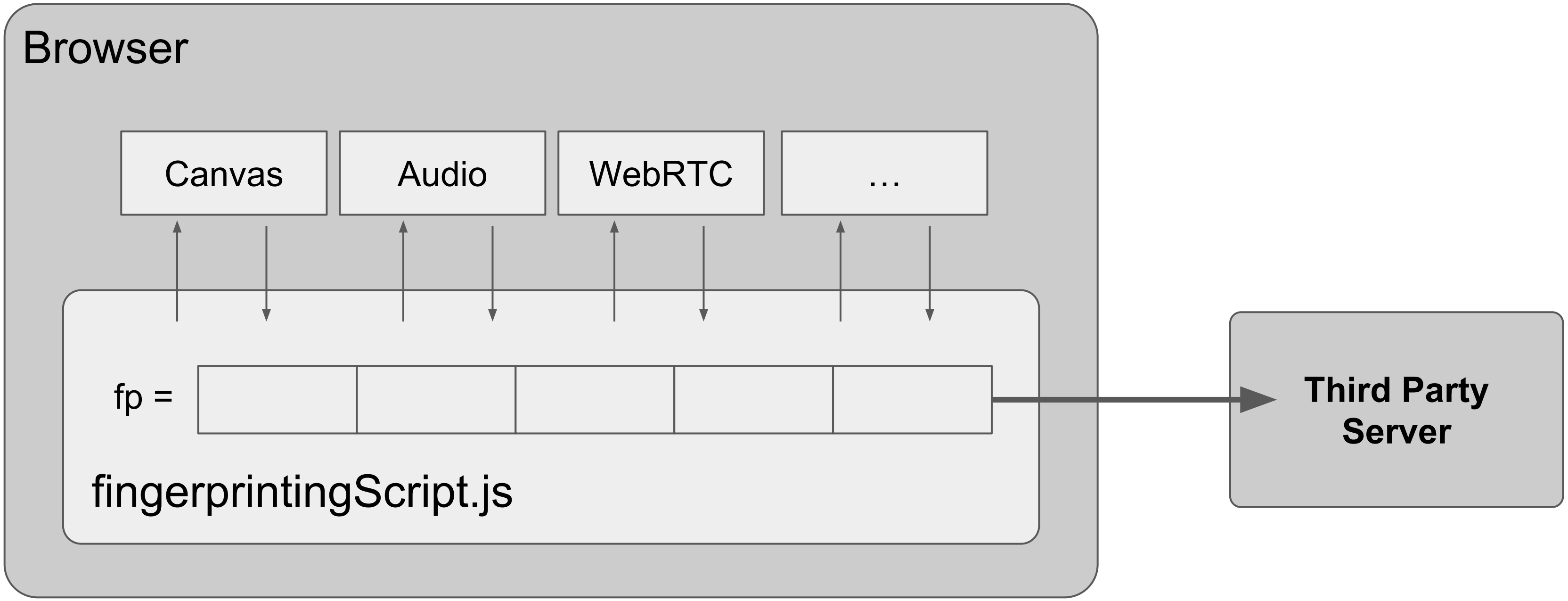
Measurement Logs

```
(SCRIPT_URL, “getContext”, “2d”)
(SCRIPT_URL, “fillText”, “hello world”, 2, 15)
(SCRIPT_URL, “fillStyle”, “#f60”)
(SCRIPT_URL, “fillRect”, 125, 1, 62, 20)
(SCRIPT_URL, “toDataURL”, “data: …” )
```

Post-measurement Analysis

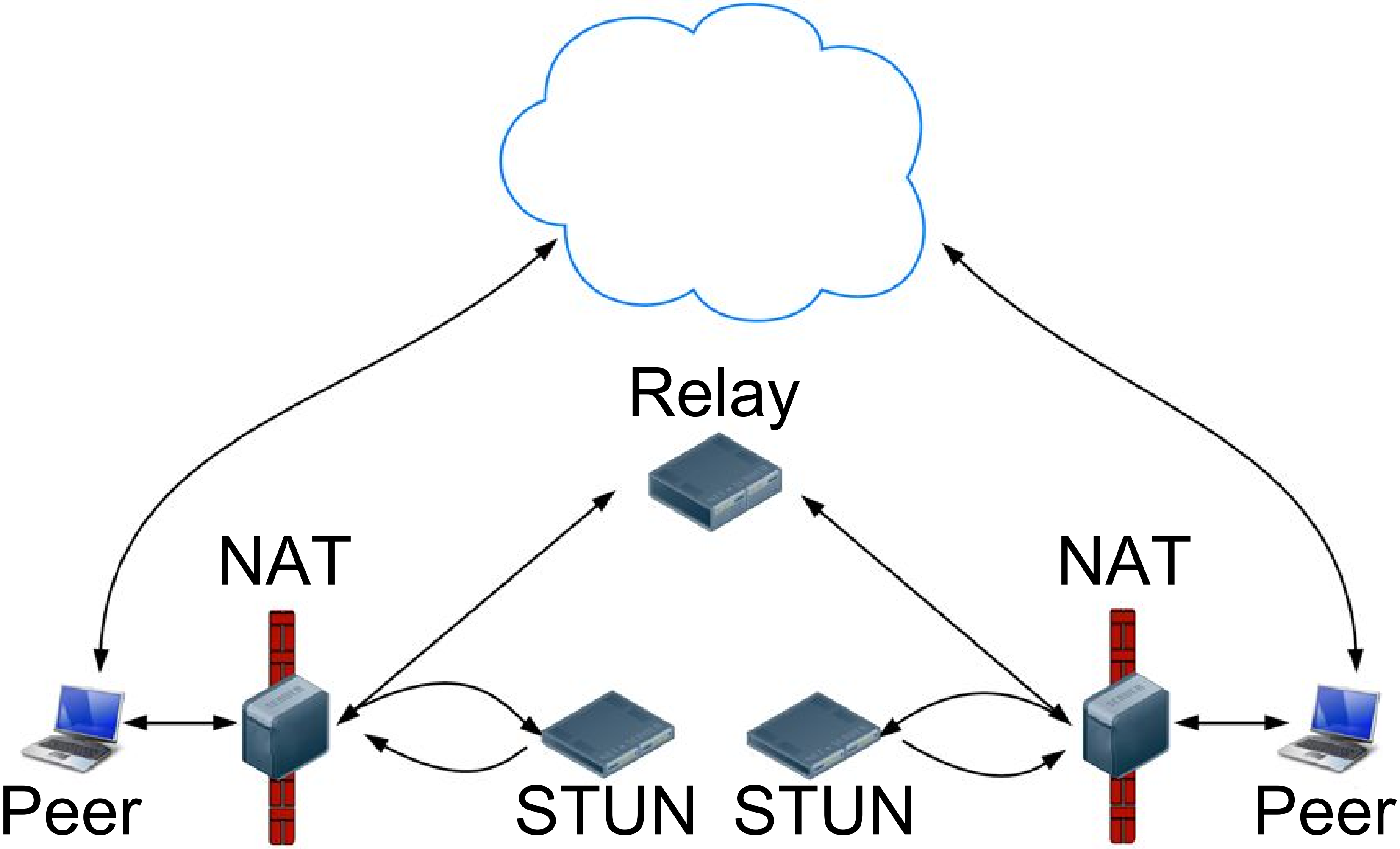
1. Examine API use for fingerprinting
2. Check for tampering / instrumentation inspection

Detecting Fingerprinting



1. Observe a sequence of API calls
2. Techniques clustered together
3. Results of calls combined and sent to server
4. Limited API use beyond that for fingerprinting

Abusing WebRTC candidate generation for tracking



Source: http://www.html5rocks.com/en/tutorials/webrtc/basics/

WebRTC `dataChannel` requires no permissions

Without user intervention, a tracking script can:

1. Reveal the user’s real IP address when behind a VPN
2. Reveal the user’s local IP address for each local interface.

WebRTC `dataChannel` requires no permissions

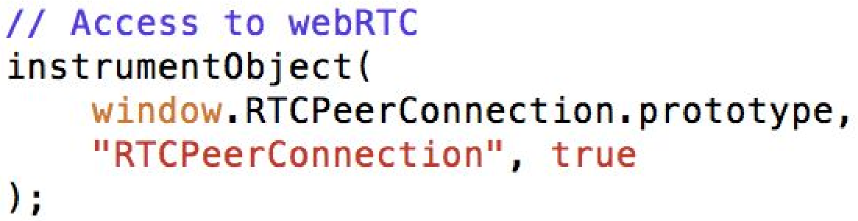
Without user intervention, a tracking script can:

1. Reveal the user’s real IP address when behind a VPN
2. Reveal the user’s local IP address for each local interface.

More identifying for corporate and university users.

Measuring the use of WebRTC for tracking

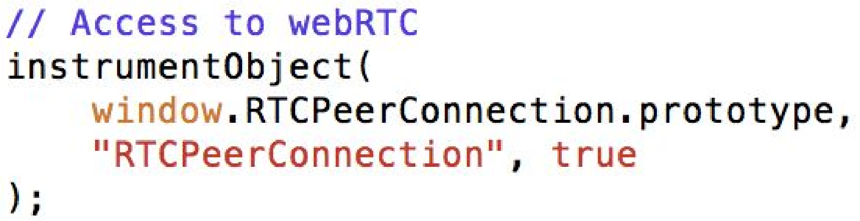
Measurement Code:



```
// Access to webRTC
instrumentObject(
    window.RTCPeerConnection.prototype,
    "RTCPeerConnection", true
);
```

Measuring the use of WebRTC for tracking

Measurement Code:



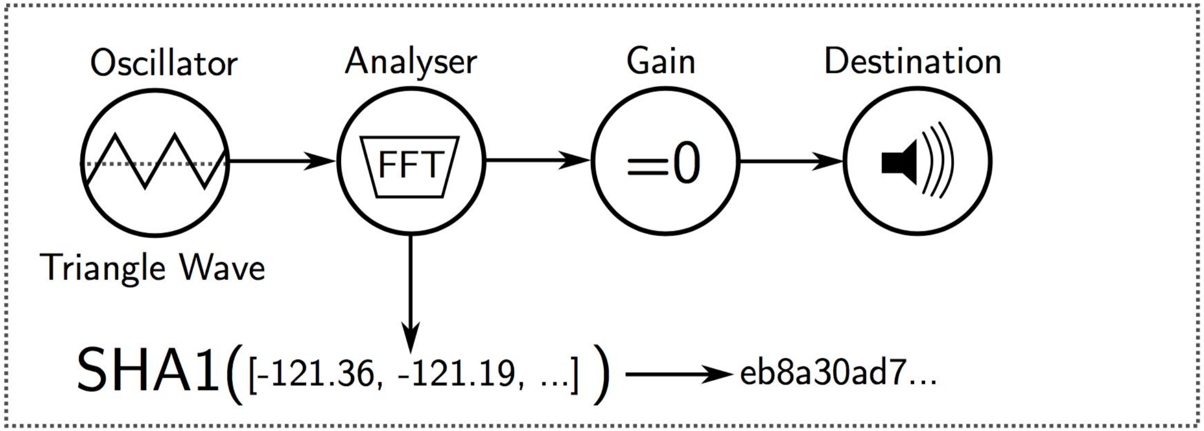
```
// Access to webRTC
instrumentObject(
    window.RTCPeerConnection.prototype,
    "RTCPeerConnection", true
);
```

~90% of unsolicited `dataChannel` use on homepages is for tracking

57 scripts on 625 sites.

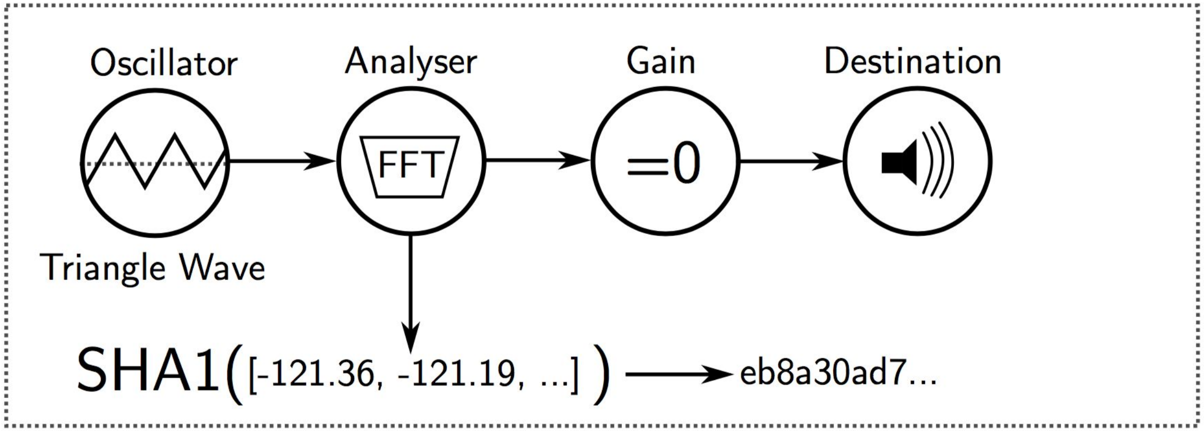
Using AudioContext for fingerprinting

Used by:
`cdn-net.com` script

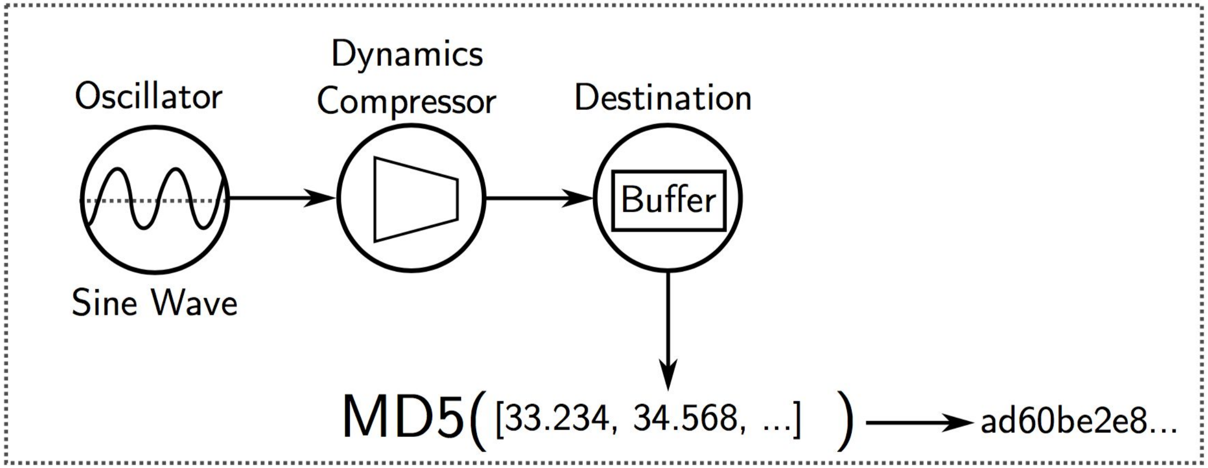


Using AudioContext for fingerprinting

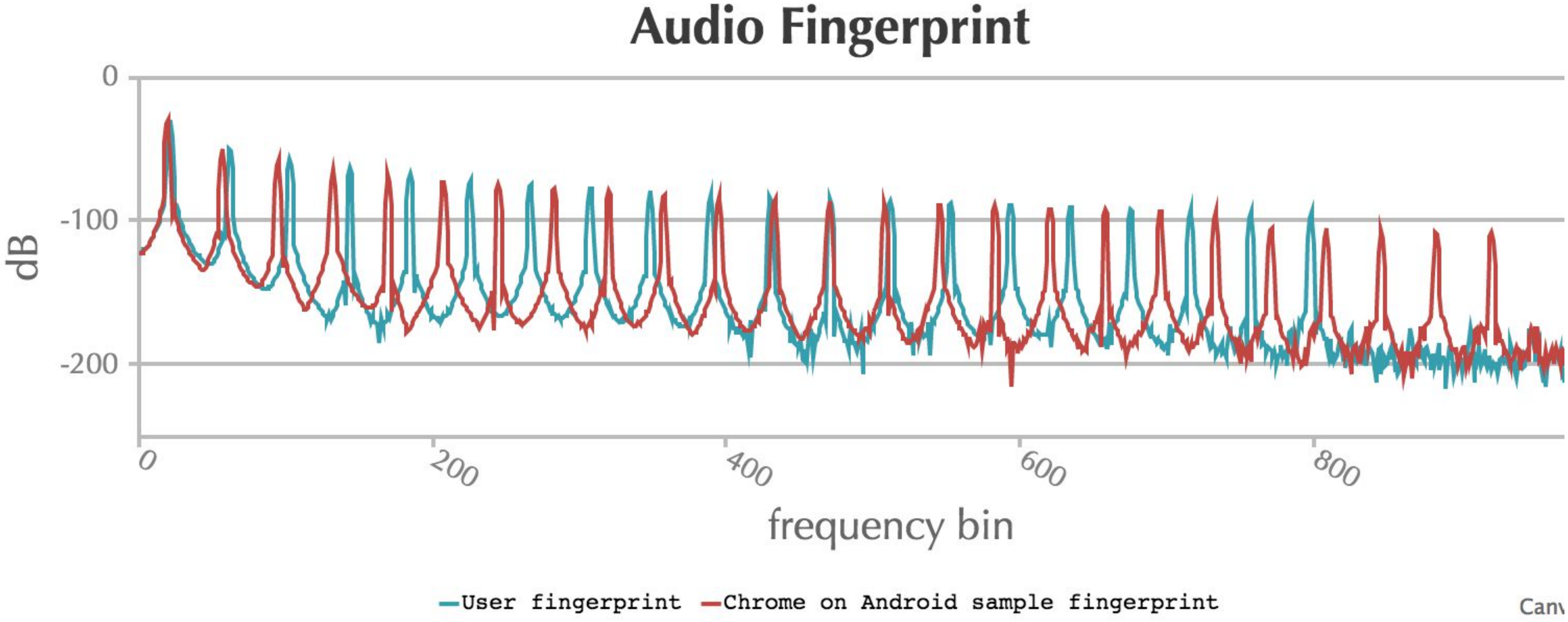
Used by:
`cdn-net.com` script



Used by:
`pxi.pub` and
`ad-score.com` scripts



Using AudioContext for fingerprinting

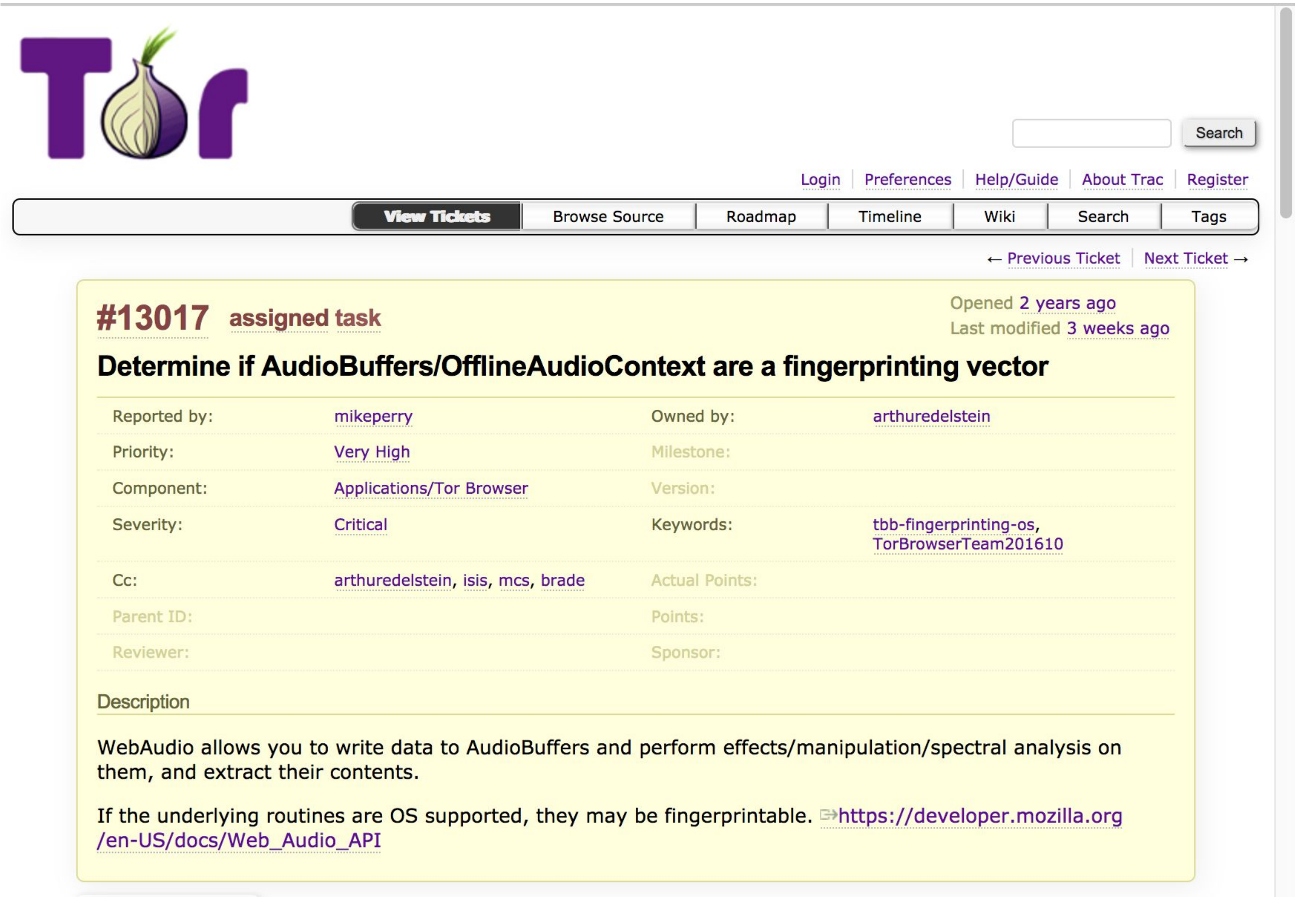


Live test page: https://audiofingerprint.openwpm.com/

Implications for Tor Browser

271 samples from the Tor Browsers

- 7 distinct fingerprints (2 fingerprints account for 80% of samples)
- Overlap with fingerprints from Firefox shows these largely reveal OS of device


Tor
Search
Login | Preferences | Help/Guide | About Trac | Register
View Tickets | Browse Source | Roadmap | Timeline | Wiki | Search | Tags
← Previous Ticket | Next Ticket →

#13017 assigned task

Opened 2 years ago
Last modified 3 weeks ago

Determine if AudioBuffers/OfflineAudioContext are a fingerprinting vector

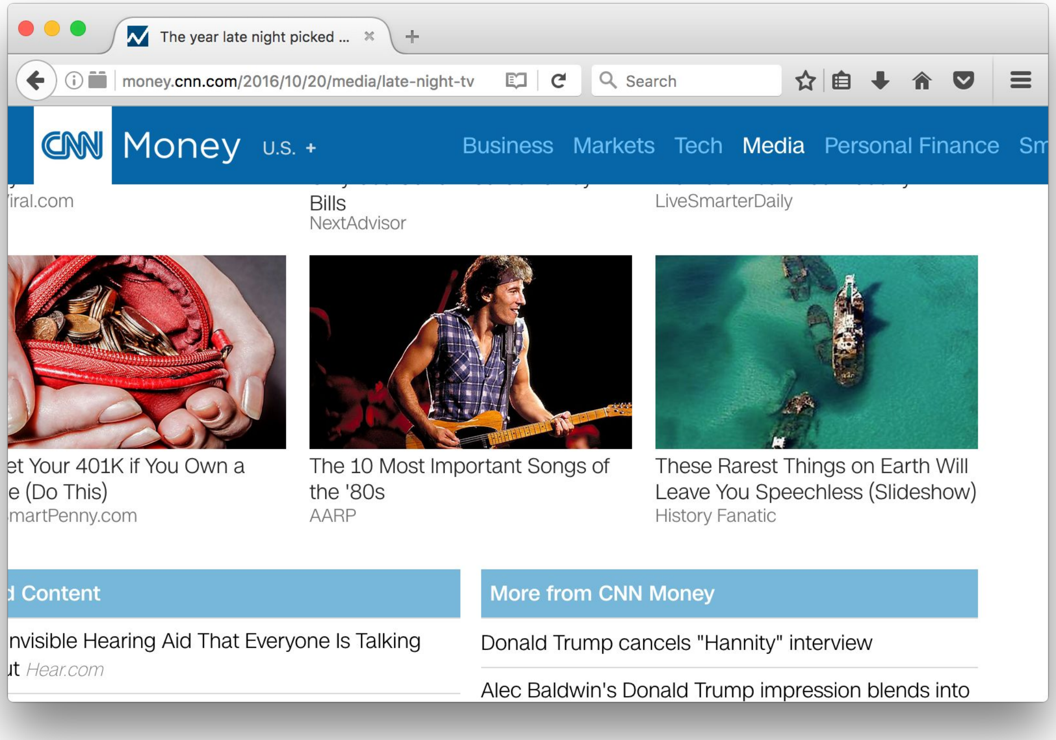
Reported by:	mikeperry	Owned by:	arthuredelstein
Priority:	Very High	Milestone:	
Component:	Applications/Tor Browser	Version:	
Severity:	Critical	Keywords:	tbb-fingerprinting-os, TorBrowserTeam201610
Cc:	arthuredelstein, isis, mcs, brade	Actual Points:	
Parent ID:		Points:	
Reviewer:		Sponsor:	

Description

WebAudio allows you to write data to AudioBuffers and perform effects/manipulation/spectral analysis on them, and extract their contents.

If the underlying routines are OS supported, they may be fingerprintable. https://developer.mozilla.org/en-US/docs/Web_Audio_API

Using Battery Status to Track


The year late night picked ...
money.cnn.com/2016/10/20/media/late-night-tv
Search
CNN Money
U.S. +
Business
Markets
Tech
Media
Personal Finance
Sm

iral.com

Bills
NextAdvisor

LiveSmarterDaily

et Your 401K if You Own a
e (Do This)
martPenny.com

The 10 Most Important Songs of the '80s
AARP

These Rarest Things on Earth Will Leave You Speechless (Slideshow)
History Fanatic

d Content

nvisible Hearing Aid That Everyone Is Talking
ut *Hear.com*

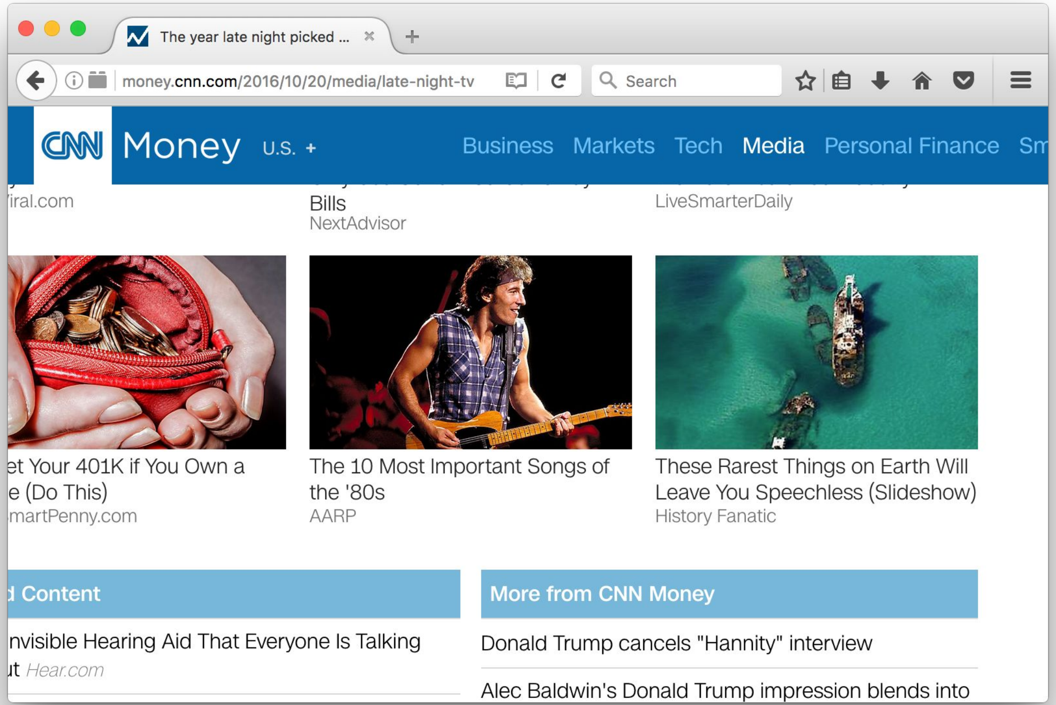
More from CNN Money

Donald Trump cancels "Hannity" interview

Alec Baldwin's Donald Trump impression blends into

The Leaking Battery, Olejnik et. al. (2015)

Using Battery Status to Track



The year late night picked ...
money.cnn.com/2016/10/20/media/late-night-tv
Search
CNN Money U.S. +
Business Markets Tech Media Personal Finance Sm

iral.com

Bills
NextAdvisor

LiveSmarterDaily

et Your 401K if You Own a
e (Do This)
martPenny.com

The 10 Most Important Songs of
the '80s
AARP

These Rarest Things on Earth Will
Leave You Speechless (Slideshow)
History Fanatic

d Content

nvisible Hearing Aid That Everyone Is Talking
ut *Hear.com*

More from CNN Money

Donald Trump cancels "Hannity" interview

Alec Baldwin's Donald Trump impression blends into

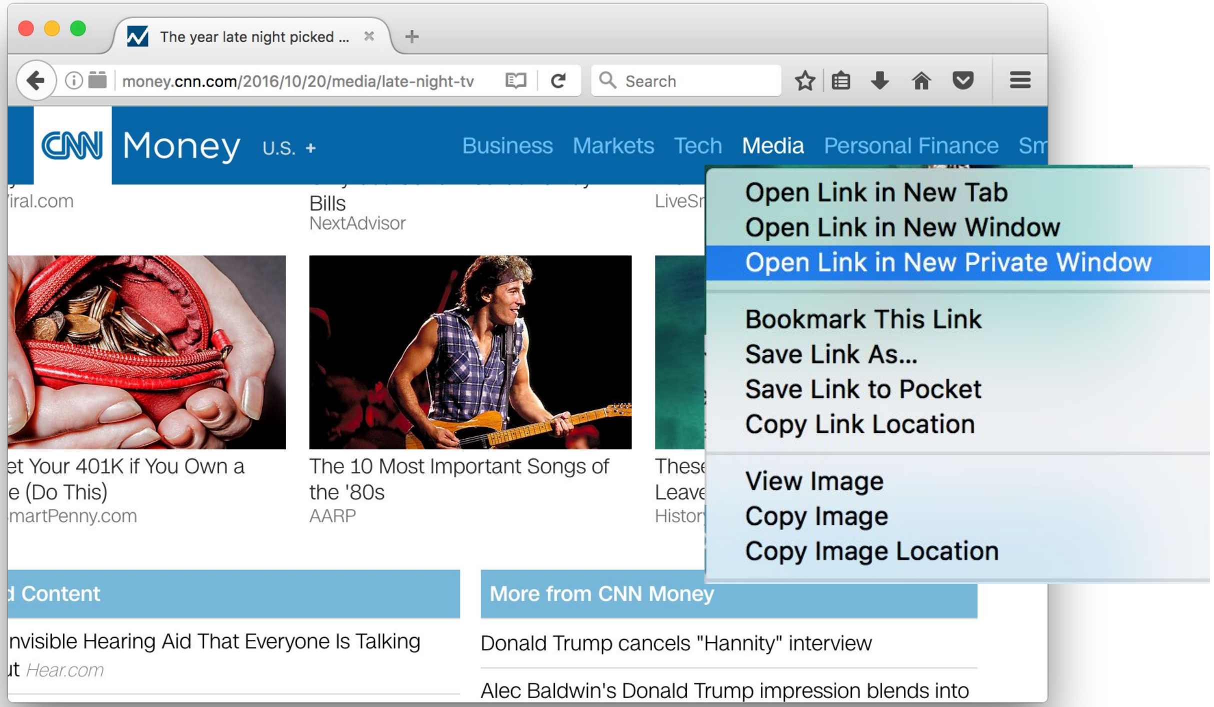
Battery Status:

`level: 0.11`

`dischargeTime: 12867`

The Leaking Battery, Olejnik et. al. (2015)

Using Battery Status to Track

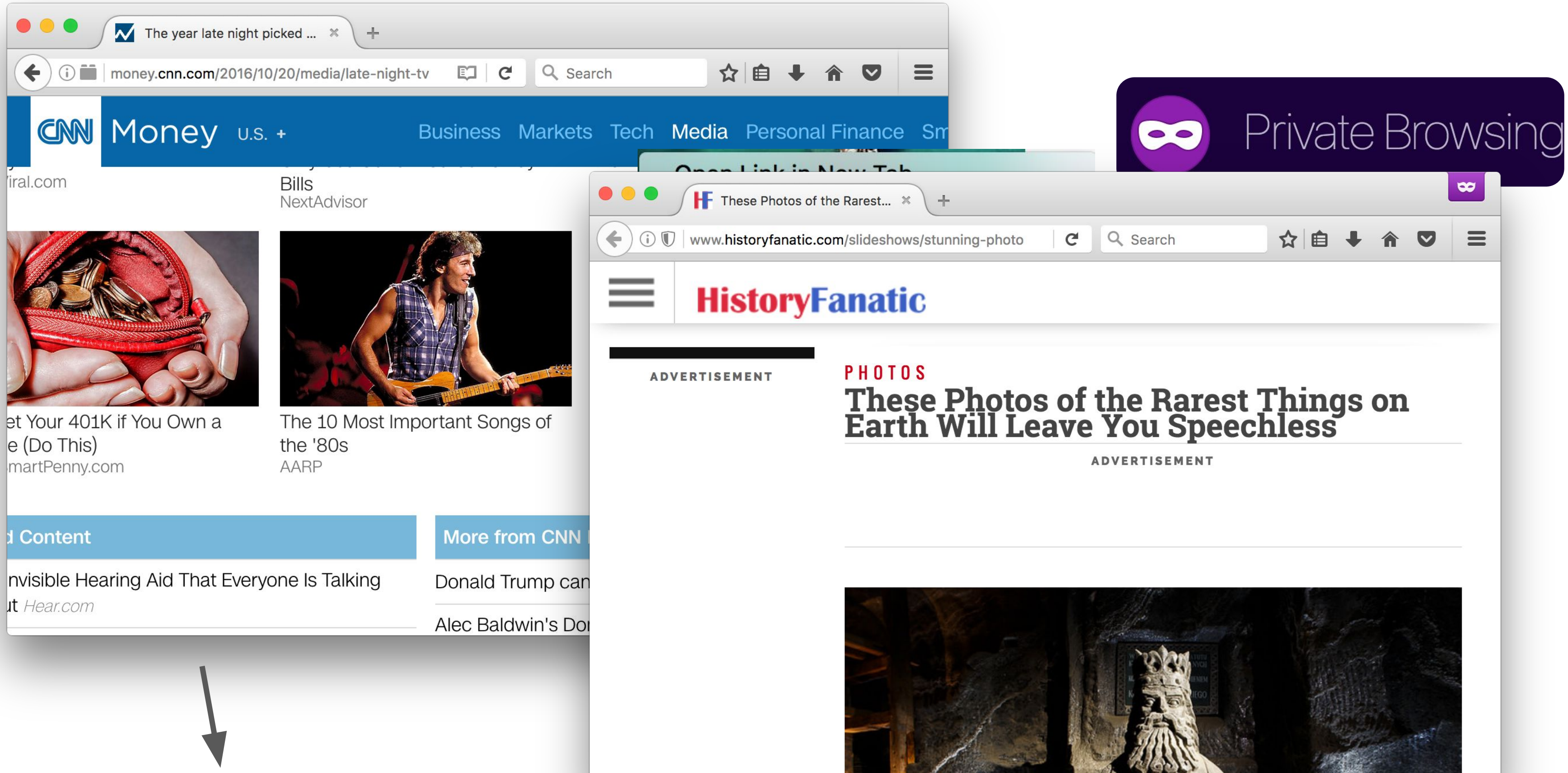


Battery Status:

```
level: 0.11
dischargeTime: 12867
```

The Leaking Battery, Olejnik et. al. (2015)

Using Battery Status to Track

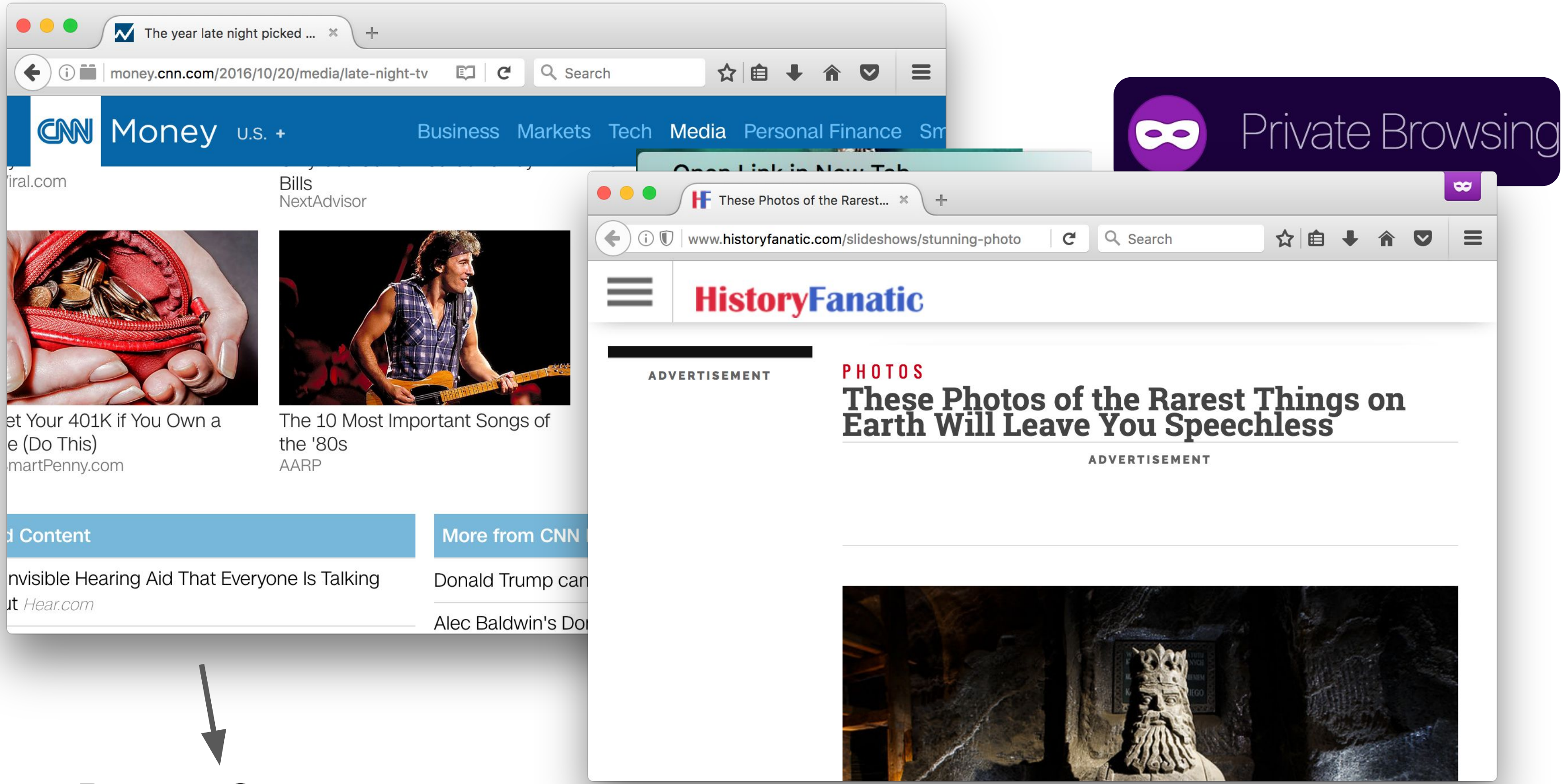

The year late night picked ...
money.cnn.com/2016/10/20/media/late-night-tv
Search
CNN Money U.S. +
Business Markets Tech Media Personal Finance Sm
iral.com
Bills
NextAdvisor
et Your 401K if You Own a
e (Do This)
martPenny.com
The 10 Most Important Songs of the '80s
AARP
d Content
More from CNN
nvisible Hearing Aid That Everyone Is Talking
ut Hear.com
Donald Trump can
Alec Baldwin's Do
Open Link in New Tab
Private Browsing
These Photos of the Rarest...
www.historyfanatic.com/slideshows/stunning-photo
Search
HistoryFanatic
ADVERTISEMENT
PHOTOS
These Photos of the Rarest Things on Earth Will Leave You Speechless
ADVERTISEMENT

Battery Status:

```
level: 0.11
dischargeTime: 12867
```

The Leaking Battery, Olejnik et. al. (2015)

Using Battery Status to Track

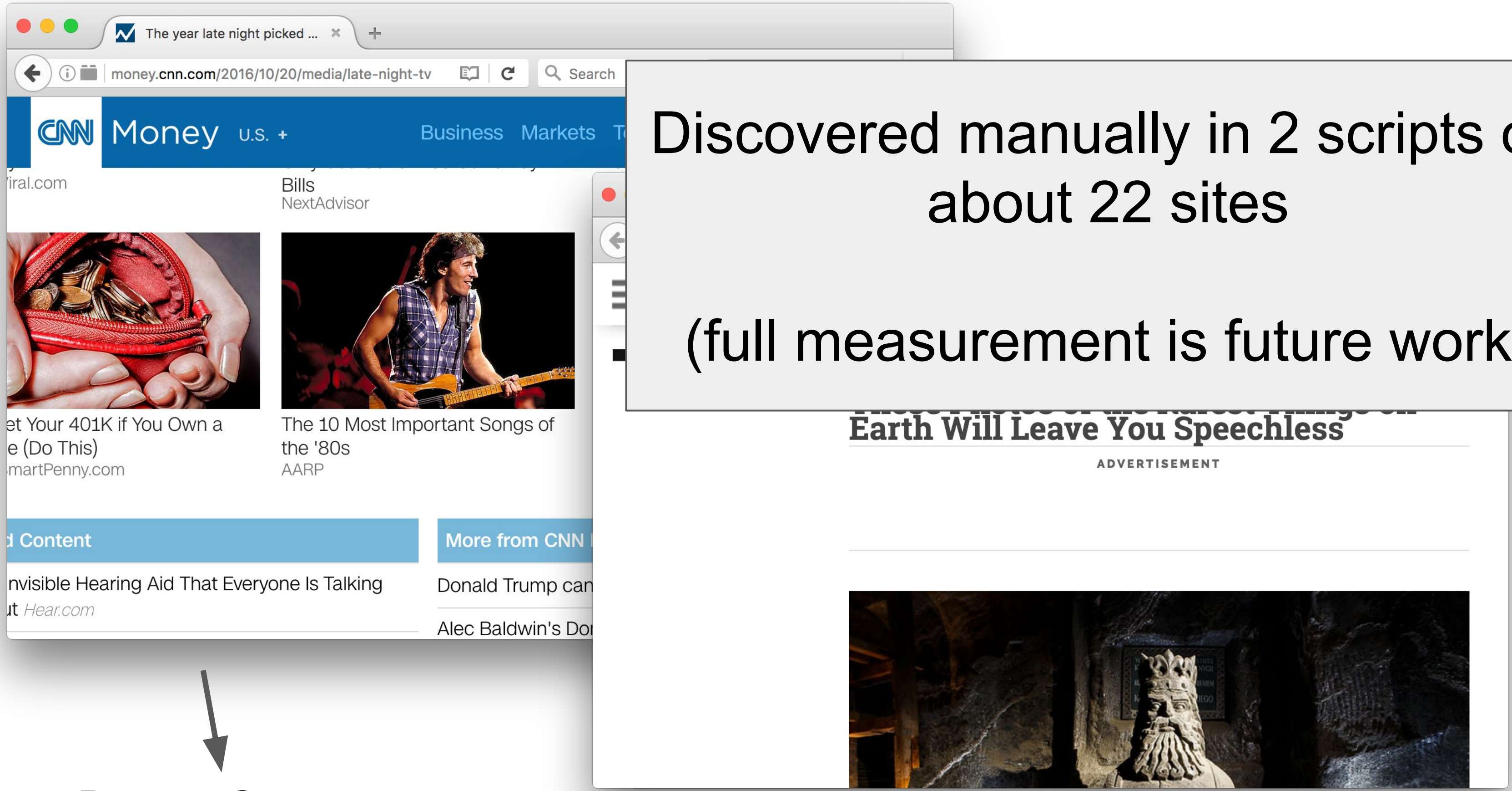

The year late night picked ...
money.cnn.com/2016/10/20/media/late-night-tv
Search
CNN Money U.S. +
Business Markets Tech Media Personal Finance Sm
Bills
NextAdvisor
et Your 401K if You Own a
e (Do This)
martPenny.com
The 10 Most Important Songs of
the '80s
AARP
d Content
More from CNN
nvisible Hearing Aid That Everyone Is Talking
ut Hear.com
Donald Trump can
Alec Baldwin's Do
Open Link in New Tab
Private Browsing
These Photos of the Rarest...
www.historyfanatic.com/slideshows/stunning-photo
Search
HistoryFanatic
ADVERTISEMENT
PHOTOS
These Photos of the Rarest Things on Earth Will Leave You Speechless
ADVERTISEMENT

Battery Status:
`level: 0.11`
`dischargeTime: 12867`

Battery Status:
`level: 0.11`
`dischargeTime: 12867`

The Leaking Battery, Olejnik et. al. (2015)

Using Battery Status to Track



The year late night picked ...
money.cnn.com/2016/10/20/media/late-night-tv
Search
CNN Money U.S. +
Business Markets T

iral.com

Bills
NextAdvisor

et Your 401K if You Own a
e (Do This)
martPenny.com

The 10 Most Important Songs of the '80s
AARP

d Content

nvisible Hearing Aid That Everyone Is Talking
ut Hear.com

More from CNN

Donald Trump can

Alec Baldwin's Do

Earth Will Leave You Speechless

ADVERTISEMENT

Discovered manually in 2 scripts on about 22 sites

(full measurement is future work)

Battery Status:
```
level: 0.11
dischargeTime: 12867
```

Battery Status:
```
level: 0.11
dischargeTime: 12867
```

The Leaking Battery, Olejnik et. al. (2015)

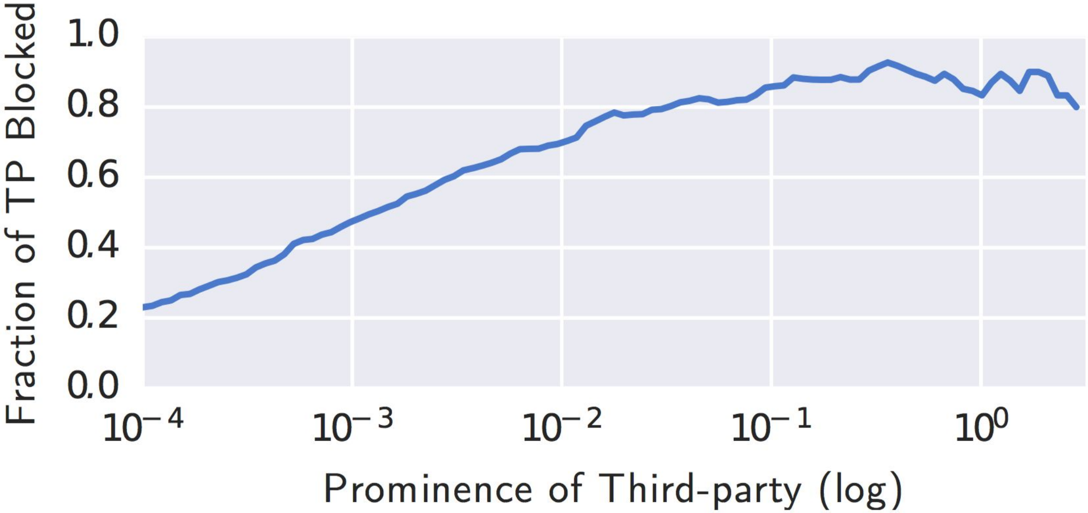
Do Privacy Tools Help?

Privacy tools effectively block stateful tracking

- Third-party cookie blocking
  - 32 out of 50,000 sites work around blocking by redirecting the top-level domain
  - Average number of third-parties per site reduced from ~18 to ~13
- Ghostery
  - Average number of third-parties per site reduced from ~18 to ~3
  - Very few third-party cookies are set

Privacy tools effectively block stateful tracking

- Third-party cookie blocking
  - 32 out of 50,000 sites work around blocking by redirecting the top-level domain
  - Average number of third-parties per site reduced from ~18 to ~13
- Ghostery
  - Average number of third-parties per site reduced from ~18 to ~3
  - Very few third-party cookies are set



Crowdsourced lists miss fingerprinters

	EasyList + EasyPrivacy	
Technique	Percentage of Scripts	Percentage of Sites

Crowdsourced lists miss fingerprinters

	EasyList + EasyPrivacy	
Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%

Crowdsourced lists miss fingerprinters

EasyList + EasyPrivacy

Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%
Canvas Font	10%	91%

Crowdsourced lists miss fingerprinters

EasyList + EasyPrivacy

Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%
Canvas Font	10%	91%
WebRTC	5%	6%

Crowdsourced lists miss fingerprinters

	EasyList + EasyPrivacy	
Technique	Percentage of Scripts	Percentage of Sites
Canvas	25%	88%
Canvas Font	10%	91%
WebRTC	5%	6%
AudioContext	6%	2%

1. Our measurement platform

2. Insights from our 1-million-site measurement

3. **Next steps**

Repeated measurements are needed

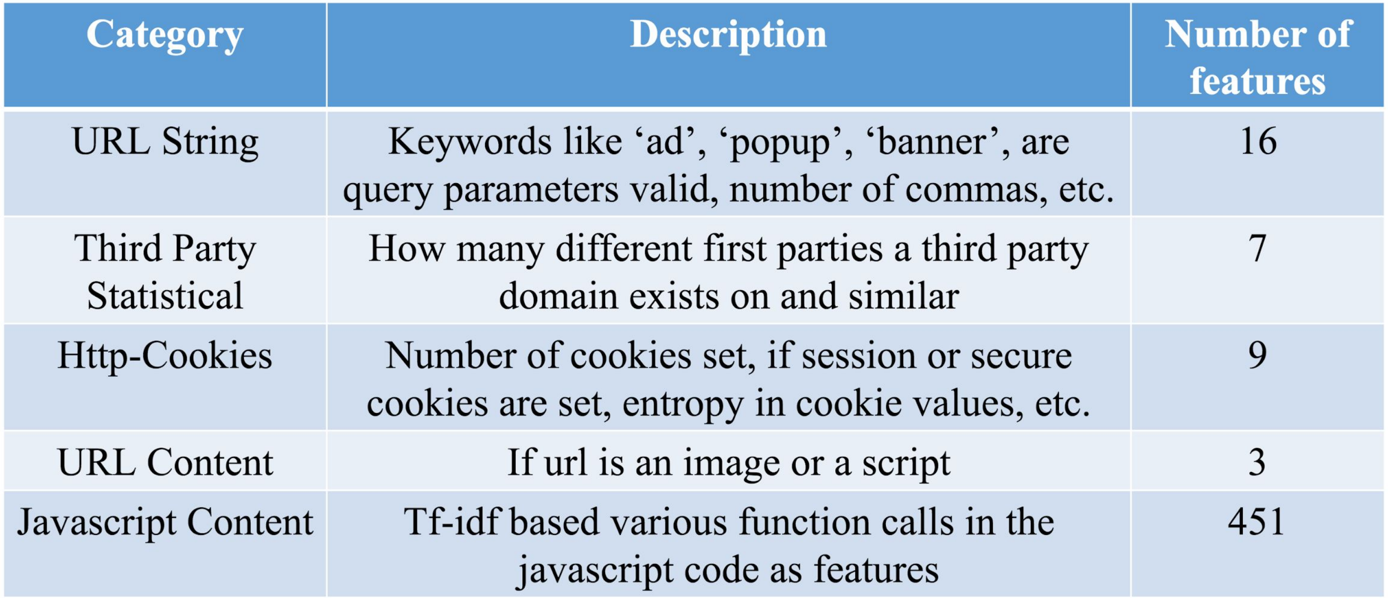
Use of canvas fingerprinting over time:

May 2014: 5% of the top 100k sites

Aug 2014: ~0.1% of the top 100k sites

Jan 2016: 2.6% of the top 100k sites

Machine learning to detect fingerprinters



Category	Description	Number of features
URL String	Keywords like ‘ad’, ‘popup’, ‘banner’, are query parameters valid, number of commas, etc.	16
Third Party Statistical	How many different first parties a third party domain exists on and similar	7
Http-Cookies	Number of cookies set, if session or secure cookies are set, entropy in cookie values, etc.	9
URL Content	If url is an image or a script	3
Javascript Content	Tf-idf based various function calls in the javascript code as features	451

- Monthly, 1-million-site view of the web
- Benefit from extensive instrumentation of OpenWPM

Master’s Thesis: *Using Machine Learning for Online Tracking Protection and Ad Blocking* by Shivam Agarwal

Takeaways

1. Trackers are employing an increasingly diverse set of techniques
2. Measurement heavily influences and controls the adoption of new techniques and tracking norms.
3. Crowdsourced tracking protection misses less popular trackers/techniques
4. Frequent measurement and automated detection provide a path forward

Takeaways

Thanks for listening!

1. Trackers are employing an increasingly diverse set of techniques
2. Measurement heavily influences and controls the adoption of new techniques and tracking norms.
3. Crowdsourced tracking protection misses less popular trackers/techniques
4. Frequent measurement and automated detection provide a path forward

Full Paper: senglehardt.com/papers/ccs16_online_tracking.pdf

Email: ste@cs.princeton.edu **Twitter:** @s_englehardt **Web:** senglehardt.com

Image Assets from the Noun Project:
Database by Creative Stall; programmer by Hadi Davodpour