Brave, Fingerprinting and Privacy on the Web

Me (the early years)

- Grew up in Chicago ...actual Chicago
- Law school, then freelance web design Started: Anchorage, AK Ended: Judge Judy Show invitation
- PhD in Computer Science
 University of Illinois at Chicago

Me, now

- **Privacy Researcher at Brave** Research to improve privacy in the browser
- Co-Chair of PING

Privacy reviews of new web standards

Academic Collaborator
 "Pure" research

Brave in a Slide

- Privacy focused
- Alternative funding model for the web
- Research and engineering focused
- Browsers and infrastructure now, more to come...



Overview

- 1. Why websites track (and how much)
- 2. "Classic" tracking
- 3. Fingerprinting / "passive tracking"
- 4. Fingerprinting counter measures
- 5. Anti-finger printing exercise
- 6. Privacy protections in Brave
- 7. Wrapping up

Overview

- 1. Why websites track (and how much)
- 2. "Classic" tracking
- 3. Fingerprinting / "passive tracking"
- 4. Fingerprinting counter measures
- 5. Anti-finger printing exercise
- 6. Privacy protections in Brave
- 7. Wrapping up

Why Does Tracking Exist?









Welcome the The "First" Banner Ad

Yes, this site is supposed to look this way. After all, this is what most web pages looked like back on October 27, 1994 -- the day that Wired Magazine flipped the switch on its first website, hotwired.com, starting a revolution in web content and advertising that still reverberates today.

This site is dedicated to showing off one of the ads that ran on that site. No, it wasn't the "first" as there were a handful of other ads that ran on various sections of hotwired.com. This site is also here to tell the story of how that ad came to be, how it succeeded beyond anything we had imagined, and how we tried to set an example for how corporations could communicate with their audiences.

This site launched on October 27, 2014. It is being constantly updated, so please check back again soon for more. In the meantime, get started by clicking your mouse in the banner ad above explore these other options:







The World's Worst Website

Gratuitous use of frames is a common mistake of web designers.

Many older browsers do not support frames. They disrupt the flow of the website and can be difficult to anticipate where a page may appear when a link is clicked. <u>Click here</u> for an example of a frames page which is opening in the wrong window. Use your browser's 'Back' button to escape.

Check out these links to websites whose opinions about frames is self evident:

The "I Hate Frames" Frames
Page

Another I Hate Frames Page

The International I Hate Frames <u>Club</u>

Why Frames Suck (Most of the Time)

Angelfire Build your own FREE website at Angelfire.com Share: del.icio.us | digg | reddit | Twitter | facebook



Welcome to the World's Worst Website!

This web was designed to graphically demonstrate the most common mistakes made by new Web Page designers.

	۰,	é.		°,		
			•			
•						
		•		۰	÷ .	
٩	3	9		8	۹	
•	٥		4	•		-

· · · · · · ·

* • * • • •

• •

Where am I and where are the links to other pages?

An easy to use navigation structure is essential to any well designed website! Important information should never be more than 2 clicks away.



As you can see, this text is difficult to read. There needs to more contrast between the background color and the text color. Here's another example of a poor choice of a background/ text color and size.

Keep your backgrounds simple. White or light colors usually work best. Your background should not compete with the content of the page for the users attention. If you would like to use a background picture, select a picture that uses muted colors or format your picture as a watermark. Select text colors which will contrast well with the background picture.



Constantly running animations can be distracting when used excessively.



o o





The World's Worst Website



\$\$\$

¢¢¢







The World's Worst Website



Chicago Tribune







Identify "expensive" people here

Pay a little to advertise to them here



DISPLAY LUMAscape





But how much...

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 automated version of a full-fledged consumer browser. It supports parallelism for speed and scale, automatic recovery from failures of the underlying browser, and comprehensive browser instrumentation. We demonstrate our platform's strength in enabling researchers to rapidly detect, quantify, and characterize emerging online tracking behaviors.

1. INTRODUCTION

Web privacy measurement — observing websites and services to detect, characterize and quantify privacy-impacting behaviors — has repeatedly forced companies to improve

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 controlled by the user and breaks frequently if used for largescale measurements. Browsers themselves tend to suffer memory leaks over long sessions. In addition, *instrumenting* the browser to collect a variety of data for later analysis presents formidable challenges. For full coverage, we've found it necessary to have three separate measurement points: a network proxy, a browser extension, and a disk state monitor. Further, we must link data collected from these disparate points into a uniform schema, duplicating much of the browser's own internal logic in parsing traffic.

A large-scale view of web tracking and privacy.







But how much... a lot / too much

Overview

- 1. Why websites track
- 2. "Classic" tracking
- 3. Fingerprinting / "passive tracking"
- 4. Fingerprinting counter measures
- 5. Anti-finger printing exercise
- 6. Privacy protections in Brave
- 7. Wrapping up



Javascript

- DOM / Initial Web API
- Netscape
- Firefox
- Brave + BAT





good-site.com







good-site.com



good-site.com



Birth of the Tracking

Problem

- Authentication?
- Can't log in every time
- HTTP auth is terrible and limited
- Solution
 - Server gives token to user
 - User returns it on requests
 - Aka "cookies"



good-site.com










Web 0.0









But in the meantime...

cat-cuties.com





kozy-kittens.com



cat-cuties.com





kozy-kittens.com









cookies, cookies everywhere...

Cookies + 3p Resources

Tracking







Site B

Tracking Site











Tracking Patient Zero

- The internets "original sin"
 - cross origin resources
 - 3p cookies
 - or both...
- "I invitent Javascript and 3p script, and I've been making up for it ever sense..." (paraphrase)



"Ever-Cookies"

- Some browsers started fighting back Brave, Safari, Firefox, extensions...
- Trackers fought back Moving IDs information out of cookies, to other location

• Long list of locations

- Local and Session Storage
- HSTS
- Cache (etags, Cache API, etc)
- Plugins
- many many many more...

Overview

- 1. Why websites track (and how much)
- 2. "Classic" tracking

3. Fingerprinting / "passive tracking"

- 4. Fingerprinting in web standards
- 5. Fingerprinting counter measures
- 6. Anti-finger printing exercise
- 7. Wrapping up

Fingerprinting, what's diff?

Classic tracking

- Website stores an id on the client
- The client returns the id to the server (cookie or JS)
- The id is what allows re-identification

Fingerprinting / passive tracking

- Website finds things different about each visitor
- That difference allows re-identification

Fingerprinting, how

- Large number of semi-identifiers
 - Browser size
 - Extra fonts
 - Audio hardware
 - Video hardware
 - Installed plugins
 - Color depth
 - etc etc ...









All browser users: 3 billion people











Succeeding at Fingerprinting

- 1. **Breath of fingerprints** Large number of semi-identifiers
- 2. **Depth of fingerprints** How uniquely each identifier can... identify

Breath (examples)

- User agent string
- Installed fonts
- Canvas / WebGL
- Hardware (many)
- Height / width

User Agent String

- History of the Browser user-agent string
 https://webaim.org/blog/user-agent-string-history/
- Katamari-Damacy of identifiers

• Brave / Chrome

Mozilla/5.0 (Macintosh; Intel Mac OS X 10_15_0) AppleWebKit/537.36 (KHTML, like Gecko) Chrome/78.0.3904.50 Safari/537.36

• Easy to extract

- navigator.userAgent
- User-Agent:

Installed Fonts

Three categories of fonts

- System
- Local
- Web
- "Local" is the tricky part
 - Office
 - Photoshop
 - Goofery

Easy to extract

- plugins
- css + span + width

Example



Example



Canvas / WebGL

- Pixel Perfect: Fingerprinting Canvas in HTML5
 Keaton Mowery and Hovav Shacham
- Drawling APIs
 e.g. Drawing lines / shapes
- Standardized, but subtle differences
- Easy to extract
 - Create canvas
 - Do some drawing
 - toDataURL()





Hovav Shacham

The Geometry of Innocent Flesh on the Bone: Return-into-libc without Function Calls
Hardware Identifiers

Many Web APIs leak capabilities

- number of cores (HTML)
- number of audio channels (Web Audio API)
- num shaders and similar (WebGL API)
- device memory (Device Memory API)
- network (WebRTC, Network status API)

Semi identifying

• Easy to extract

- All browsers have subset of the above
- Most platforms have no permissions

Height / Width



Height / Width

- What does it mean?
- How to extract w/ JavaScript?
- How to extract w/o JavaScript?
- Brb 5 min (Go go go go go!)

Fingerprinting Depth



https://panopticlick.eff.org/

Fingerprinting in Practice

- Needs to be in a database...
- Hash each endpoint
- Hash each value into a single identifier...
- Nice implication: "poisionability"...

Exercise

- Read fingerprint2.js
- List as many finger-printing approaches as possible
- Understand how they're carried out
- Predict which are most identifying

Overview

- 1. Why websites track (and how much)
- 2. "Classic" tracking
- 3. Fingerprinting / "passive tracking"

4. Fingerprinting counter measures

- 5. Anti-finger printing exercise
- 6. Privacy protections in Brave
- 7. Wrapping up

Fingerprinting Countermeasures

- Remove the functionality
- Make the functionality consistent
- Restrict access (permissions, 1p vs 3p, user gesture, etc)
- Noise
- "Privacy budget"

Remove the functionality

- Delete JS end point
- Remove the HTTP header
- Remove the runtime capability

Consistency

- Make every browser return the same value
- ... or, most?
- Not that diff in practice from "removing"

Restrict access

- Permission prompt
- User gesture
- 1p vs. 3p
- "Site engagement"

Noise

- Stenography
- Make different every time

Original With Hidden Data



Privacy Budget

- Allow some identifiability
- After "identifiability budget" is exhausted do... something
- Google folks love it
- Everyone else is... skeptical

Overview

- 1. Why websites track (and how much)
- 2. "Classic" tracking
- 3. Fingerprinting / "passive tracking"
- 4. Fingerprinting counter measures
- 5. Anti-finger printing exercise
- 6. Privacy protections in Brave
- 7. Wrapping up

Fingerprint2 Again...

- Choose two fingerprinting vectors to combat
 - Propose counter measures
- Choose two fingerprinting vectors that are hard
 - Why are counter measures hard?

Fingerprint2 pt 3...

- Pretend your the attacker
- How would you respond to those defenses...

Fingerprint2 pt 4...

- Pretend your the defender again
- How would you modify your defenses given the previous round...

Overview

- 1. Why websites track (and how much)
- 2. "Classic" tracking
- 3. Fingerprinting / "passive tracking"
- 4. Fingerprinting counter measures
- 5. Anti-finger printing exercise
- 6. Privacy protections in Brave
- 7. Wrapping up



Shields UP for this site If a site appears broken, try shields down	Shields
japantoday.com 11 Items blocked	Global protection
 V 11 Cross-site trackers blocked O Connections upgraded to HTTPS 	from tracking
 ✓ 0 Scripts blocked Cross-site cookies blocked 	 On by default
 ✓ 0 Cross-site device recognition blocked ✓ Simple View Change global shield defaults 	 Can be disabled if needed



- Block cross site trackers
- Lists of known tracking websites
- Refuse to load
- Both community and Brave generated

Blocking Cross-Site Trackers in Brave

- EasyList and EasyPrivacy Used by AdBlock Plus, etc.
- Disconnect
 Used by Firefox, extensions
- uBlock Origin
 Excellent blocking extension
- Brave generated
 Öpen source, shared with community



 Don't send identifiers to third party sites

- Send to "main" site
- Same with other storage methods





Brave Blocks Tracking Cookies





Tracking Site









- Reduce finger printing vectors
- Currently:
 - Hardware identifiers
 - Canvas
 - WebGL
 - \circ Audio
- Planned:
 - Fonts
 - User agent
 - \circ Screen size

Under Exploration Possible Privacy Protections

- Restrictions on third-party scripts
- Identifying tracking behaviors, not just scripts / URLs
- Query parameters filtering
- Bounce tracking
- Much more...

Overview

- 1. Why websites track (and how much)
- 2. "Classic" tracking
- 3. Fingerprinting / "passive tracking"
- 4. Fingerprinting counter measures
- 5. Anti-finger printing exercise
- 6. Privacy protections in Brave

7. Wrapping up

Unasked for Advice

- Brave is hiring, keep us in mind
- Privacy is more than just web, there's lots to do
- Don't accept privacy as a feature...
- Choose your employer with values in mind

Thanks!

Pete Snyder

Privacy Researcher pes@brave.com @pes10k

- Questions?
 - Standards work?
 - Privacy jobs?
 - Brave business model
 - BAT / Block chain
 - Anything else?

