Privacy

Michalis Polychronakis

Stony Brook University
Privacy

“The right of an entity (normally a person), acting in its own behalf, to determine the degree to which it will interact with its environment, including the degree to which the entity is willing to share information about itself with others.”  [RFC2828]

Beyond private data (messages/files):

- Activities (browsing history, daily routine, voice commands, …)
- Location (3/4G, GPS, WiFi, cameras, …)
- Preferences (“likes,” Amazon, Netflix, …)
- Health (Fitbit, iWatch, …)

…
Real-world Privacy

Large-scale data collection examples
- Credit cards, Metrocards, loyalty cards
- Street/public space cameras, tolls, badge readers
- Named tickets (travel, events, services)
  ...

Part of our everyday activities and personal information is (voluntarily or compulsorily) recorded
- Information from different sources can be correlated

Did you buy your Metrocard with your credit card?

The same happens in the online world…
Third parties have access to…

Our email  (Gmail, Yahoo, …)
Our files  (Dropbox, Google Drive, …)
Our finances  (e-banking, credit reporting, Mint, …)
Our communication  (Instant messengers, Zoom, …)
Our traffic  (WiFi hotspots, ISPs, …)
Our location  (3/4/5G, GPS, WiFi, BLE, …)
Our activities  (browsing history, daily routine, …)
Our preferences  (“Likes,” Amazon, Netflix, …)
Our health  (Fitbit, iWatch, 23andMe, …)
…
Hacking of Government Computers Exposed 21.5 Million People

BY JULIE HIRSCHFELD DAVIS  JULY 9, 2015

Katherine Archuleta, director of the Office of Personnel Management, right, at hearing before the House Oversight and Government Reform Committee last month. Mark Wilson/Getty Images

WASHINGTON — The Obama administration on Thursday revealed that 21.5 million people were swept up in a colossal breach of government computer systems that was far more damaging than officials initially reported.
Credit firm Equifax says 143m Americans' social security numbers exposed in hack

- Atlanta-based company says ‘criminals’ accessed personal data
- Before notifying public, Equifax executives sold $1.8m in shares

Credit monitoring company Equifax says a breach exposed the social security numbers and other data of about 143 million Americans.

After discovering the breach, but before notifying the public, three Equifax senior executives sold shares in the company worth almost $1.8m. Since the public announcement, the company’s share price has tumbled.

The Atlanta-based company said Thursday that “criminals” exploited a US website application to access files between mid-May and July of this year.

It said consumers’ names, social security numbers, birth dates, addresses and, in some cases, driver’s license numbers were exposed. Credit card numbers for about 209,000 US consumers were also accessed.

“This is clearly a disappointing event for our company, and one that strikes at the heart of who we are and what we do,” said the company’s chairman and
World's Biggest Data Breaches & Hacks

Selected events over 30,000 records

UPDATED: Jan 2021

size: records lost

Facebook 420,000,000
Indian citizens 275,000,000
Microsoft 250,000,000
OxyData 380,000,000
Twitter 330,000,000

Marriott International 383,000,000
Capital One 100,000,000

Canva 130,000,000
Apollo 200,000,000
Aadhaar

Chirbox 100,000,000
Chinese resume leak 202,000,000
Firebase 50,000,000

Spambot 711,000,000
Nametests 120,000,000
Quora 100,000,000

Twitter 330,000,000

Armed With Facebook 'Likes' Alone, Researchers Can Tell Your Race, Gender, and Sexual Orientation

REBECCA J. ROSEN | MAR 12 2013, 2:59 PM ET

But the deeper aspects of your personality remain hard to detect.
Facebook knows you’re gay before you do

3/20/13 4:29pm by Jon Green 39 Comments

Am I the only one creeped out that Facebook is now guessing, sometimes correctly, if its users are gay?

In the world of Big Data, our private lives are increasingly becoming intermingled with the shadowy, yet public, world of cyberspace.

Whenever we go online we are providing data that can be used to market to us; from Google searches to Facebook likes to eBay purchases, we are inputting data into a series of mathematical models which make incredibly educated guesses about the kinds of people we are.

Facebook creepily offers help to a gay guy thinking of “coming out”

Enter Matt. As BuzzFeed notes, Matt was your typical Facebook user who suddenly found an ad in his news feed for help in coming out. The weird thing was that Matt “did” need help coming out, and understandably he was more than a bit curious as to how Facebook knew.

At first, Matt wondered if Facebook had accessed his text messages, as he had confided in a close friend the previous
Facebook users unwittingly revealing intimate secrets, study finds

Personal information including sexuality and drug use can be correctly inferred from public 'like' updates, according to study.
How Smart TVs in Millions of U.S. Homes Track More Than What’s On Tonight

By Sapna Maheshwari
July 5, 2018

The growing concern over online data and user privacy has been focused on tech giants like Facebook and devices like smartphones. But people’s data is also increasingly being vacuumed right out of their living rooms via their televisions, sometimes without their knowledge.

In recent years, data companies have harnessed new technology to immediately identify what people are watching on internet-connected TVs, then using that information to send targeted advertisements to other devices in their homes. Marketers, forever hungry to get their products in front of the people most likely to
China Is Using Facial Recognition Technology to Send Jaywalkers Fines Through Text Messages

It’s the latest update to a widely deployed facial recognition surveillance system in China.

By Daniel Oberhaus

March 28, 2018, 8:00am

In China, law enforcement agencies have been using advanced biometric technology to track citizens for years. These technologies are part of a coordinated national effort to create the “omnipresent, completely

MORE LIKE THIS
How Target Figured Out A Teen Girl Was Pregnant Before Her Father Did

Every time you go shopping, you share intimate details about your consumption patterns with retailers. And many of those retailers are studying those details to figure out what you like, what you need, and which coupons are most likely to make you happy. Target, for example, has figured out how to data-mine its way into your womb, to figure out whether you have a baby on the way long before you need to start buying diapers.

Charles Duhigg outlines in the New York Times how Target tries to hook would-be customers at that crucial moment before they turn into rampant — and profitable — spenders...
Web Browsing Tracking

Webpages are often mashups of content loaded from different sources
- Ads, images, videos, widgets, …
- IMG URLs, IFRAMEs, JavaScript, web fonts, social widgets, …
- Hosted on third-party servers: CDNs, cloud providers, ad networks, …

A third party involved in many different websites can track user visits across all those websites
- Multiple third parties may collude to expand their collective “view”

Trackers want to learn two key pieces of information
- What webpage was visited
- Who visited it
Microsoft Announces Continuum: Turning Windows 10 Phones Into Desktop

Microsoft just demonstrated one of the intriguing possibilities from its single platform/multiple device strategy. Instead of just putting a Windows 10 phone on a Windows 10 PC, it can run a full Windows 10 desktop on it, including apps and devices.

The main app here is called Disconnect, which is a privacy tool that lets you see which sites are tracking you, and disallow them.

Disconnect works by monitoring the web pages you visit and connecting them to a data model. If the site is identified as tracking you, Disconnect will block it, and put it in its list of blacklisted sites.

The Disconnect app shows a graph of the sites you visit, and allows you to block or unblock them. It also provides a list of known tracking sites, so you can easily add them to the blacklist.
What webpage was visited?

HTTP Referer [sic] header

The full URL of the webpage from which a link was followed
Useful for statistics/analytics, bad for privacy
Can be turned off through browser options/extensions
HTML5 rel="noreferrer" anchor attribute to indicate to the user agent not to send a referrer when following the link
Most browsers have started sending only the origin part in cross-origin requests

Page-specific, session-specific, user-specific URLs

Unique URL per page (even for the same resource) ➔ track what page was visited
Unique URL per session/user ➔ distinguish between visits from different users
Firefox 87 trims HTTP Referrers by default to protect user privacy

Dimi Lee and Christoph Kerschbaumer  |  March 22, 2021

We are pleased to announce that Firefox 87 will introduce a stricter, more privacy-preserving default Referrer Policy. From now on, by default, Firefox will trim path and query string information from referrer headers to prevent sites from accidentally leaking sensitive user data.

Referrer headers and Referrer Policy

Browsers send the HTTP Referrer header (note: original specification name is ‘HTTP Referer’) to signal to a website which location “referred” the user to that website’s server. More precisely, browsers have traditionally sent the full URL of the referring document (typically the URL in the address bar) in the HTTP Referrer header with virtually every navigation or subresource (image, style, script) request. Websites can use referrer information for many fairly innocent uses, including analytics, logging, or for optimizing caching.

Unfortunately, the HTTP Referrer header often contains private user data: it can reveal which articles a user is reading on the referring website, or even include information on a user’s account on a website.
Tracking URLs are also commonly used in promotional emails

**Embedded image loading**

This is an active email address! Detect the time a user viewed a message
The request reveals much more: user agent, device, location, …

**Embedded links**

Learn which email addresses resulted in visits (click-through rate)

**Default behavior of email clients varies**

Gmail used to block images by default, now uses image proxy servers
Tracking through unique images still possible: senders can track the first time a message is opened (user’s IP is not exposed though)
Who visited the page?

**Browsing to a web page reveals a wealth of information**

**Source IP address**
- Not very accurate (e.g., NAT, DHCP, on-the-go users) but still useful

**Third-party cookies:** precise user tracking
- Easy to block (configurable in most browsers, defaults vary)

“**Evercookies:**” exploit alternative browser state mechanisms
- ETags, HTML5 session/local/global storage, plugin-specific storage, …

**Browser/device fingerprinting:** recognize unique system characteristics
- Browser user agent, capabilities, plugins/extensions, system fonts, screen resolution, time zone, and numerous other properties
<table>
<thead>
<tr>
<th></th>
<th>Brave</th>
<th>Chrome</th>
<th>Edge</th>
<th>Firefox</th>
<th>Safari</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Mechanism</strong></td>
<td>Shields</td>
<td>n/a</td>
<td>Tracking prevention</td>
<td>Enhanced Tracking Protection (ETP)</td>
<td>Intelligent Tracking Prevention (ITP)</td>
</tr>
<tr>
<td><strong>Deployed in</strong></td>
<td>0.55.18</td>
<td>n/a</td>
<td>78.0.276.8</td>
<td>69.0</td>
<td>Safari 11</td>
</tr>
<tr>
<td><strong>Default protection mode</strong></td>
<td>Default Shield settings</td>
<td>n/a</td>
<td>Balanced</td>
<td>Standard</td>
<td>ITP enabled</td>
</tr>
<tr>
<td><strong>Classification of “known trackers”</strong></td>
<td>Multiple filter lists</td>
<td>n/a</td>
<td></td>
<td>Trust Protection Lists (with engagement and organization mitigation)</td>
<td></td>
</tr>
<tr>
<td><strong>Cookies in 3rd party context</strong></td>
<td>Restrict access in subresource requests, Partitioned access in frame, Partitioned access for known trackers</td>
<td>No restrictions.</td>
<td>Access restricted for known trackers.</td>
<td>Access restricted for known trackers.</td>
<td>All access restricted, except with Storage Access API.</td>
</tr>
</tbody>
</table>
How to Read Your Report

You will see a summary of your overall tracking protection. The first section gives you a general idea of what your browser configuration is blocking (or not blocking). Below that is a list of specific browser characteristics in the format that a tracker would view them. We also provide descriptions of how they are incorporated into your fingerprint.

How Can Trackers Track You?

Trackers use a variety of methods to identify and track users. Most often, this includes tracking cookies, but it can also include browser fingerprinting. Fingerprinting is a sneaker way to track users and makes it harder for users to regain control of their browsers. This report measures how easily trackers might be able to fingerprint your browser.

How Can I Use My Results to Be More Anonymous?

Knowing how easily identifiable you are, or whether you are currently blocking trackers, can help you know what to do next to protect your privacy. While most trackers are based on widely available information, your own browser fingerprint will contain additional information.

Is Your Browser:

<p>| | |</p>
<table>
<thead>
<tr>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>Blocking tracking ads?</td>
<td>Yes</td>
</tr>
<tr>
<td>Blocking invisible trackers?</td>
<td>Yes</td>
</tr>
<tr>
<td>Protecting you from fingerprinting?</td>
<td>Your browser has a unique fingerprint</td>
</tr>
</tbody>
</table>

Note: Because tracking techniques are complex, subtle, and constantly evolving, Cover Your Tracks does not measure all forms of tracking and protection.

Your Results

Your browser fingerprint appears to be unique among the 273,004 tested in the past 45 days.

Currently, we estimate that your browser has a fingerprint that conveys at least 18.06 bits of identifying information.

The measurements we used to obtain this result are listed below. You can read more about our methodology, statistical results, and some defenses against fingerprinting here.
What do web tracking techniques really track?

Distinguish between different visitors

*Track anonymous individuals*

Actually: track the pages visited by a particular browser running on a particular device

Better: distinguish between different *persons*

*Track named individuals*

The transition is easy…

Personally identifiable information (PII) is often voluntarily provided to websites:

- Social networks, cloud services, web sites requiring user registration, …
- Cookies/sessions are associated with PII

Contamination: trackers may collude with services

- Previously “anonymous” cookies/fingerprints can be associated with named individuals
VERIZON’S ‘PERMA-COOKIE’ IS A PRIVACY-KILLING MACHINE
AT&T charges $29 more for gigabit fiber that doesn’t watch your Web browsing

AT&T goes head to head against Google in KC on fiber and targeted ads.

by Jon Brodkin - Feb 16, 2015 12:36pm EST

AT&T’s gigabit fiber-to-the-home service has just arrived in Kansas City, and the price is the same as Google Fiber—if you let AT&T track your Web browsing history.
T-Mobile will sell your web-usage data to advertisers unless you opt out

Data sales begin April 26 unless you opt out; T-Mobile claims it'll be anonymous.

JON BRODKIN - 3/9/2021, 5:35 PM

"Starting April 26, 2021, T-Mobile will begin a new program that uses some data we have about you, including information we learn from your web and device usage data (like the apps installed on your device) and interactions with our products and services for our own and 3rd party advertising, unless you tell us not to," T-Mobile said in a privacy notice. "When we share this information with third parties, it is not tied to your name or information that directly identifies you."

For directions on how to opt out of the expanded data sharing, see the first section of the T-Mobile privacy notice. We've heard from customers who say they've had problems opting out so you may have to try multiple links or make multiple attempts. There's another list of opt-out links here and a link here to change the "Do Not Sell" setting. "T-Mobile will not sell personal data to third parties when you tell us not to," the company's privacy notice said.
Users register on trackers!

Social widgets are prevalent

- 2.8+ billion Facebook (monthly active) users
- Twitter, LinkedIn, Pinterest, AddThis, …
- OS/app integration

A growing part of our browsing history can be tracked by social networking services

- Not as merely anonymous visitors, but as named persons
- Just visiting the page is enough (no interaction needed)
- Cross-device tracking
### Social Sharing Usage Distribution in the Top 1 Million Sites

<table>
<thead>
<tr>
<th>Technology</th>
<th>Websites</th>
<th>%</th>
</tr>
</thead>
<tbody>
<tr>
<td>AddThis</td>
<td>61,528</td>
<td>6.15</td>
</tr>
<tr>
<td>Facebook Like</td>
<td>37,506</td>
<td>3.75</td>
</tr>
<tr>
<td>ShareThis</td>
<td>29,531</td>
<td>2.95</td>
</tr>
<tr>
<td>Add to Any</td>
<td>21,180</td>
<td>2.12</td>
</tr>
<tr>
<td>Reddit</td>
<td>20,724</td>
<td>2.07</td>
</tr>
<tr>
<td>Baidu Share</td>
<td>8,567</td>
<td>0.86</td>
</tr>
<tr>
<td>Bitmx24</td>
<td>6,861</td>
<td>0.69</td>
</tr>
<tr>
<td>Yotpo</td>
<td>4,232</td>
<td>0.42</td>
</tr>
<tr>
<td>Sassy Social Share</td>
<td>4,225</td>
<td>0.42</td>
</tr>
<tr>
<td>POWr</td>
<td>2,882</td>
<td>0.29</td>
</tr>
<tr>
<td>Instagram API</td>
<td>2,791</td>
<td>0.28</td>
</tr>
<tr>
<td>Sina Weibo</td>
<td>2,531</td>
<td>0.25</td>
</tr>
<tr>
<td>Facebook Embedded Posts</td>
<td>1,834</td>
<td>0.18</td>
</tr>
<tr>
<td>Juicer</td>
<td>1,818</td>
<td>0.18</td>
</tr>
<tr>
<td>bShare</td>
<td>1,608</td>
<td>0.16</td>
</tr>
</tbody>
</table>

Other 19%

<table>
<thead>
<tr>
<th>Social Sharing Technology</th>
<th>Percentage</th>
</tr>
</thead>
<tbody>
<tr>
<td>AddThis</td>
<td>27%</td>
</tr>
<tr>
<td>Bitmx24</td>
<td>3%</td>
</tr>
<tr>
<td>Baidu Share</td>
<td>4%</td>
</tr>
<tr>
<td>Reddit</td>
<td>9%</td>
</tr>
<tr>
<td>Add to Any</td>
<td>9%</td>
</tr>
<tr>
<td>ShareThis</td>
<td>13%</td>
</tr>
<tr>
<td>Facebook Like</td>
<td>16%</td>
</tr>
<tr>
<td>Instagram API</td>
<td>2%</td>
</tr>
<tr>
<td>Sina Weibo</td>
<td>2%</td>
</tr>
<tr>
<td>Facebook Embedded Posts</td>
<td>2%</td>
</tr>
<tr>
<td>Juicer</td>
<td>2%</td>
</tr>
<tr>
<td>bShare</td>
<td>2%</td>
</tr>
<tr>
<td>Other</td>
<td>19%</td>
</tr>
</tbody>
</table>
Existing Solutions

Log out
   Some cookies persist
Block third-party cookies
   Not always effective
Block social widgets completely
Incognito mode
All existing solutions disable content personalization
   Privacy vs. functionality dilemma
First Party Isolation (Firefox)

AKA Cross-Origin Identifier Unlinkability (Tor Browser)

All identifier sources and browser state are scoped (isolated) using the URL bar domain

Cookies, cache, HTTP Authentication, DOM Storage, Flash cookies, SSL and TLS session resumption, HSTS and HPKP supercookies, OCSP, …

Example: `tracker.com` sets/reads cookies in `bbc.com` and `cnn.com`

Before: `tracker.com` can track the same person on both sites

After: `tracker.com` will see two different cookies

Third party cookies are stored with a tag of the first party (e.g., `bbc.com.tracker.com` and `cnn.com.tracker.com`)
How does Facebook Container work?

The Facebook "Like" and "Share" buttons that appear on shopping, news and other sites contain Facebook trackers. Even if you don't use them, Facebook uses these buttons to track you. Facebook Container blocks these trackers and will display a fence icon to show you where these trackers were removed.

When you visit Facebook, the add-on loads it in another tab and the fence icon is displayed in your address bar. This puts Facebook in its own boundary with other Facebook-owned sites, including Instagram and Messenger. You can allow other sites into the Facebook Container boundary, but this will allow Facebook to track more of your web activity.

When you visit a non-Facebook site that has Facebook trackers, Facebook Container will alert you and block these trackers.

You can add a website to Facebook Container if you prefer to allow Facebook to see your activity on that site.
Single Sign-on/Social Login

Pros

Convenience: fewer passwords to remember
Rich experience through social features
Outsource user registration and management

Cons

Same credentials for multiple sites
User tracking
Access to user’s profile
Take it or leave it
Location Tracking

IP addresses reveal approximate location information

MaxMind statistics: 99.8% accurate on a country level, 90% accurate on a state level in the US, and 81% accurate for cities in the US within a 50 kilometer radius

Mobile devices allow for precise location tracking

Cell tower triangulation/trilateration

GPS, GLONASS, …

WiFi access points in known locations

Per-app permissions

Android vs. iOS:
installation vs. usage time
Attention, Shoppers: Store Is Tracking Your Cell

By STEPHANIE CLIFFORD and QUENTIN HARDY    JULY 14, 2013

Like dozens of other brick-and-mortar retailers, Nordstrom wanted to learn more about its customers — how many came through the doors, how many were repeat visitors — the kind of information that e-commerce sites like Amazon have in spades. So last fall the company started testing new technology that allowed it to track customers’ movements by following the Wi-Fi signals from their smartphones.

But when Nordstrom posted a sign telling customers it was tracking them, shoppers were unnerved.

“We did hear some complaints,” said Tara Darrow, a spokeswoman for the store. Nordstrom ended the experiment in May, she said, in part because of the comments.

Nordstrom’s experiment is part of a movement by retailers to gather data about in-store shoppers’ behavior and moods, using video surveillance and signals from their cellphones and apps to learn.
Fitness tracking app Strava gives away location of secret US army bases

Data about exercise routes shared online by soldiers can be used to pinpoint overseas facilities

Sensitive information about the location and staffing of military bases and spy outposts around the world has been revealed by a fitness tracking company.

The details were released by Strava in a data visualisation map that shows all the activity tracked by users of its app, which allows people to record their exercise and share it with others.

The map, released in November 2017, shows every single activity ever uploaded to Strava – more than 3 trillion individual GPS data points, according to the company. The app can be used on various devices including smartphones and fitness trackers like Fitbit to see popular running routes in major cities, or spot individuals in more remote areas who have unusual exercise patterns.

However, over the weekend military analysts noticed that the map is also

Strava suggests military users 'opt out' of heatmap as row deepens

Read more
Privacy and Security Built In

AirTag is designed from the ground up to keep location data private and secure. No location data or location history is physically stored inside AirTag. Communication with the Find My network is end-to-end encrypted so that only the owner of a device has access to its location data, and no one, including Apple, knows the identity or location of any device that helped find it.

AirTag is also designed with a set of proactive features that discourage unwanted tracking, an industry first. Bluetooth signal identifiers transmitted by AirTag rotate frequently to prevent unwanted location tracking. iOS devices can also detect an AirTag that isn’t with its owner, and notify the user if an unknown AirTag is seen to be traveling with them from place to place over time. And even if users don’t have an iOS device, an AirTag separated from its owner for an extended period of time will play a sound when moved to draw attention to it. If a user detects an unknown AirTag, they can tap it with their iPhone or NFC-capable device and instructions will guide them to disable the unknown AirTag.
Online Behavioral Tracking

Many of our daily activities are being recorded

- What we are interested in (Searches, Likes, …)
- What we read (News, magazines, blogs, …)
- What we buy (Amazon, Freshdirect, …)
- What we watch (Netflix, Hulu, …)
- What we eat (Seamless, GrubHub, …)
- Where we eat (Opentable, Foursquare, …)
- Where we go (online travel/hotel/event booking)
- What we own/owe (e-banking, credit services, Mint, …)

Mobile apps make behavioral tracking easier and more accurate

Behavioral profiles have desirable and not so desirable uses

- Recommendations, content personalization, insights, …
- Targeted advertising, price discrimination (e.g., insurance premiums based on past behavior, higher prices for high-end device users), …
https://twitter.com/morgon/status/1346464501319622657
Apple privacy label
Health and Activity

Health records
   How securely are they handled and stored?

Devices track our activities and health
   Activity tracking devices
   Health monitoring devices
   Mobile phones

Many upload all data to the “cloud”…
   Who can access them?

Doctor/hospital health portals managed by third parties
Protecting Privacy

Preferably through technical means, not promises

Avoid collecting personal data in the first place
  iOS vs. Android, DuckDuckGo vs. Google, …

Block tracking, fingerprinting, profiling, …
  Brave, Firefox, Safari, ad blockers, …

Privacy-preserving protocols/mechanisms
  Differential privacy, on-device processing, content prefetching, …

Self-hosted services
  Only for geeks

Data privacy laws
  EU General Data Protection Regulation (GDPR), California Consumer Privacy Act (CCPA)