Tracked Without a Trace: Cookies, Fingerprinting, and Our Crumbling Privacy Online

Jenna Everard

Privacy in a Networked World
Spring 2022

In our increasingly networked world, privacy continues to face new threats despite its great value. Online tracking techniques have been around since the invention of the browser cookie in the early 1990's, but the ever-increasing sophistication of these techniques have made them difficult to detect and even more difficult to deter. Between an observed lack of awareness and an apparent lack of regulative legislation, it is becoming clear that much work remains to be done to elucidate the implications of these technologies. In this study, a combination of online user surveys and in-person interviews were conducted to analyze awareness, perspectives, and concerns surrounding these online tracking techniques. Results indicate that while the vast majority have some knowledge of tracking using browser cookies, this knowledge is both focused on negative connotations and limited in that it does not extend to other methods like supercookies or browser fingerprinting. Despite this, a general trend can be observed in a decrease in user comfort level with data being collected online as this data becomes more personal. Yet, while users are concerned for their privacy online, they feel as if their efforts, and that of privacy legislation, lack the ability to provide protection or prevent identification via means such as browser fingerprinting. Ultimately, resulting insights can be applied to direct the formation of a suite of educational resources and to propose an approach to legislative reform.

Introduction

The internet was intended to be a place where knowledge could be shared and consumed freely. However, in our modern, highly networked world, the internet seems to have become almost synonymous with a lack of privacy and, as such, a lack of this freedom. This invasion of privacy has come in the forms of various tracking technologies, namely cookies, supercookies, and browser fingerprinting. When these technologies were first introduced, widespread tracking mechanisms seemed like a thing of science fiction, but as we now know, time would prove this notion wrong.

A Brief History of Digital Privacy (or the lack thereof)

The concept of cookies first emerged around 1979, when the term "magic cookie" was used to describe the packet of information used in communication between two devices during standard C routines such as `fseek`. Fifteen years later in 1994, Lou Montulli – a computer scientist at Netscape Communications and the developer of, among other things, an early web browser and animated gifs – hatched a plan to make use of these magic cookies for web communications. He saw these cookies as a solution for the implementation of virtual shopping carts for one of the first e-commerce platforms, in which data would be stored on a user's computer rather than bogging down the platform's own servers (Kihn, 2018). In 1995, support for these cookies were integrated into Internet Explorer. Montulli would later be granted a patent for his cookie technology in 1998 (Montulli, 1998).

The general public, though, was not made aware of these cookies until the publication of a Financial Times article in 1996 (Jackson, 1996). What followed were two U.S. Federal Trade Commission hearings that focused on the privacy implications of cookies and the formation of the Internet Engineering Task Force (IETF) in 1997, of which Montulli served as one of the heads. The IETF proceeded to release a document entitled RFC2109 – RFC standing for Request for Comments – in which they recommended all

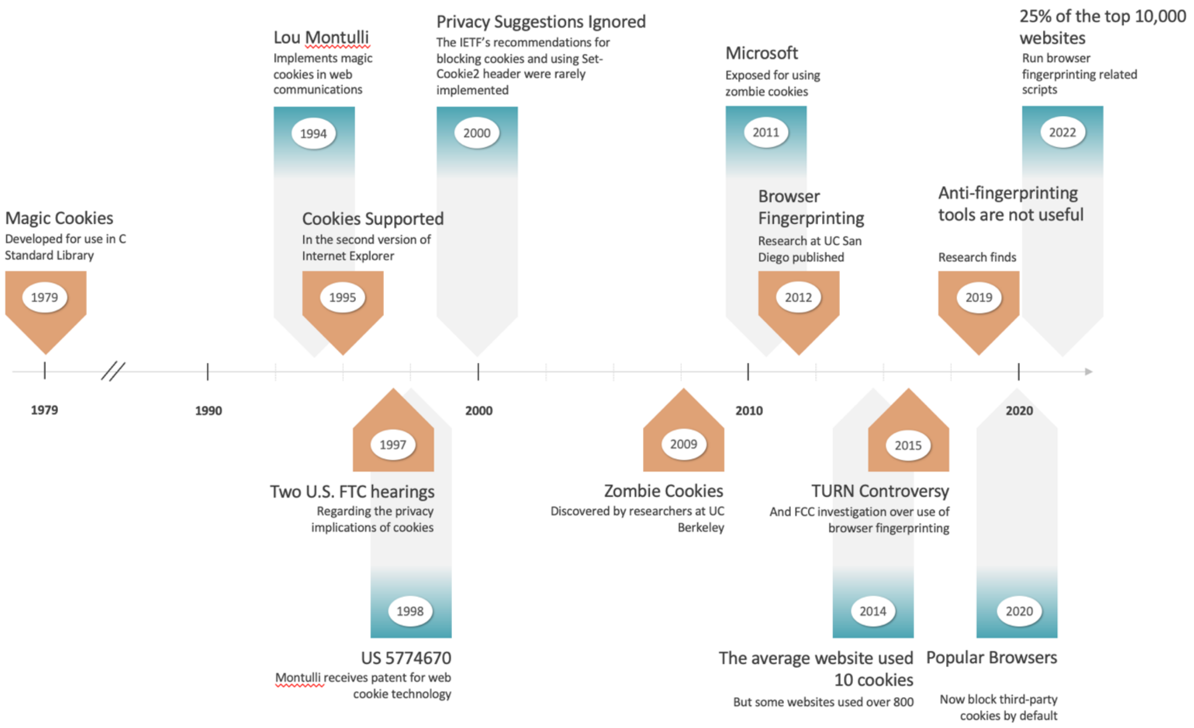

Magic Cookies
Developed for use in C Standard Library
1979
Lou Montulli
Implements magic cookies in web communications
1994
Cookies Supported
In the second version of Internet Explorer
1995
1997
Two U.S. FTC hearings
Regarding the privacy implications of cookies
1998
US 5774670
Montulli receives patent for web cookie technology
Privacy Suggestions Ignored
The IETF's recommendations for blocking cookies and using Set-Cookie2 header were rarely implemented
2000
2009
Zombie Cookies
Discovered by researchers at UC Berkeley
Microsoft
Exposed for using zombie cookies
2011
Browser Fingerprinting
Research at UC San Diego published
2012
2014
The average website used 10 cookies
But some websites used over 800
2015
TURN Controversy
And FCC investigation over use of browser fingerprinting
Anti-fingerprinting tools are not useful
Research finds
2019
2020
Popular Browsers
Now block third-party cookies by default
25% of the top 10,000 websites
Run browser fingerprinting related scripts
2022
1979
1990
2000
2010
2020

Figure 1: Historical Timeline. Shows major events in the development of browser cookies and browser fingerprinting.

third-party cookies either be blocked or automatically disabled (Kristol, 1997). This recommendation was followed by a second, RFC2965, which proposed the use of a new header that would offer improved privacy (Kristol, 2000). However, most corporations and web browsers ignored these recommendations and today, no browsers support their proposed cookie header.

With the rise in popularity of cookie-disabling technologies in the early 2000's, major privacy issues related to browser cookies became temporarily less prevalent. Derived from their intent to decrease the load on web servers, cookies were generally stored locally on a user's web browser, meaning users had a large extent of control to delete, limit, and block these cookies. However, it was around 2009 that researchers from UC Berkeley began to notice that some browser cookies, even when deleted, would keep reappearing (Soltani et. al., 2009). These aptly named "zombie cookies", also known as supercookies, presented a whole new slew of privacy concerns and controversies. In 2011, private researchers found that Microsoft was using two supercookie technologies on its website, but these were ultimately removed after public backlash. Then, in 2015, Stanford researcher Jonathan Mayer released proof that supercookies on Verizon mobile phones were being hijacked by third party advertising companies, particularly TURN, and being used to disseminate user's private information (Singer and Chen, 2015). An FCC investigation in 2016 ultimately resulted in Verizon paying a $1.35 million dollar settlement and restructuring its privacy policy (Federal Communications Commission, 2016). It did not, though, forbid Verizon from using supercookies, rather it just had to ensure user consent.

While frustrations and efforts were focused on these zombie cookies, a new, arguably more threatening, way of tracking emerged. This was browser fingerprinting. One of the landmark papers that really brought this issue to light was that of UC San Diego computer scientists Keaton Mowery and Hovav

Shacham. In 2012, they published their paper discussing their highly effective method of using HTML's canvas element to fingerprint a user's browser based on fonts, graphical renderings, etc (Mowery and Shacham, 2012). Since then, research has continued to demonstrate the accuracy with which browser fingerprints can uniquely identify an individual user and the unfortunate reality that anti-fingerprinting tools likely lack the ability to provide adequate protection.

While efforts have been made to limit the capabilities of cookies, little has been done in regard to browser fingerprinting. This past year, Apple introduced new aggressive policy protections, part of which involved explicitly notifying users when apps or websites were requesting to track them. As of 2020, most popular browsers, such as Apple Safari and Firefox, now block all third- party cookies by default. Such policies are beginning to significantly impact the revenues of companies who rely on this tracking, some estimates placing the cost for Meta at $10 billion for just this year (Conger and Chen, 2022). Browser fingerprinting, however, is absent from the majority of privacy talks and legislation. Further, there appears to be a correlation between the increase in restrictions on cookies and an increase in the use of browser fingerprinting. In 2014, it was estimated that 0.4% of the 10,000 most popular websites implemented some form of browser fingerprinting scripts. Today, this estimate has risen to over 25% (Burgess, 2022).

A "Recipe" for the Technologies

To fully comprehend the privacy implications of these technologies, it is imperative to understand the differences in which they are made and function.

The original form of web tracking, cookies are small data files that store information about us on the internet. Each time we search a specific webpage, what happens on the backend is that our browser sends an HTTP request composed specifically for that site. The web server hosting that site will then respond with an HTTP response, a chunk of code that contains necessary, legitimate information for displaying the webpage, but one which may also request further information from our browser, including cookies via a `set-cookies` header (Roesner, Kohno, & Wetherall, 2012). Once a cookie has been generated for a particular web server, it will continue to be sent with all following requests such that the web server can determine which requests are originating from the same browser, and hence the same user (Mozilla, 2022). While such tracking does serve as the foundation for the functioning of certain websites – the cookies which enable this known as first-party cookies – there are also those cookies collected by embedded web servers, such as those by advertisers. These third-party cookies can then be used to track users across many websites.

As mentioned before, the implementation of cookies as being stored on a user's local machine allows the user to have some extent of control over them. Users can choose to view their cookies, to periodically delete cookies, or to block cookies entirely! This is where the idea of a zombie cookie becomes relevant. As a colloquial term for unique identifier headers (UIDHs), zombie cookies, unlike regular cookies, are not stored on a user's machine. Rather, they are inserted into the HTTP request by an Internet Service Provider after this request leaves the user's machine and before it reaches the destination web server (Phillips, 2019). As such, users have no control over what information is stored and shared, and traditional cookie-blocking mechanisms that may be activated on a user's device are obsolete. Still, with evolving security measures, these supercookies have started to become less effective.

Enter browser fingerprinting. While cookies contain specific information about our browsers and are stored locally on our machines, a browser fingerprint is not. Not only does browser fingerprinting take place without our knowledge, it also disguises itself as necessary, legitimate queries. Websites need to collect information about our screen size,

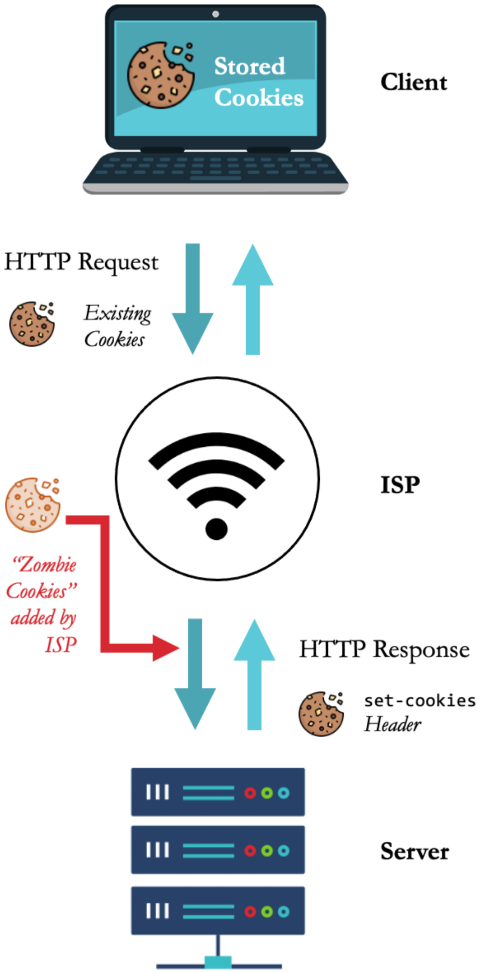


Figure 2: How Cookies Work. The red cookies and arrows provide a comparison to supercookies.

resolution, operating system, and more in order to properly display on our specific computer model. At first, it may seem like such information is not personally identifiable in any manner – and it is initially a very underdetermined system – but after collecting enough of these characteristics, these so-called browser fingerprints can actually be used to track individuals, both on a single website and across websites, browsers, and apps (Chen, 2019). Though many forms of browser fingerprinting have been established, the most widely used method, due to its high accuracy and speed, is canvas fingerprinting. In this technique, the web server uses an HTML5 canvas element to request the user's web browser to render an image with random size, text, colors, etc. (Wickramasinghe, 2021). Different devices will render this image differently depending on their graphics drivers and software, adding variables that can further constrain the fingerprinting system.

Browser fingerprinting techniques are highly accurate too! One of the papers on browser fingerprinting was a 2009 study by Jonathan Mayer, then an assistant professor in computer science at Princeton University. Developing his own browser fingerprinting strategy, he wanted to see if such seemingly insignificant differences in software and hardware from user to user could actually be used as a means of tracking. The result? It could, and with high precision too. Mayer ended up having a 96.23% success rate at tracking approximately 1,000 study participants. A year later, in 2010, Peter Eckersley of the Electronic Frontier Foundation (EFF) did a similar experiment, which he termed his Panopticlick experiment, in which he found that even with his much larger sample population of around 450,000 fingerprints, he was still able to track 83.6% (Bhagyashree, 2019). What had once been simply configuration settings for a webpage has become a unique fingerprint, non-personally identifiable data has become a means of tracking an individual, an individual who has no control over or even knowledge of what is happening.

Motivation

The privacy implications of these online tracking techniques, the lack of awareness around them, and the absence of regulations for them serve as motivation for this project. Many studies have focused on technical solutions for hindering the effectiveness of online tracking, yet little has been done with regard to user opinions and educational resources. Technical solutions only go so far, as their successful implementation and widespread usage will only result if people are aware of their need and educational information about online privacy and its importance is available and accessible. As such, this study aims to further investigate the privacy implications of cookies and browser fingerprinting and gain insights into the cognizance and perception of the average internet user. By carrying out a series of surveys and interviews, not only will we begin to answer questions surrounding user awareness and perspectives, but we will begin to spread awareness, both specifically about browser fingerprinting and more generally about how to protect our privacy in an increasingly networked world. Doing so will help initiate a collaborative thought process towards strategies, measures, and regulations that both individuals and larger communities can adapt.

Methods

Online Survey

A short, multi-part online survey study was conducted to gain information about general knowledge, perspectives, and impressions of online privacy and tracking mechanisms. Recruitment of participants consisted of two rounds. The first round was carried out from March 31st, 2022, to April 6th, 2022, and primarily focused on recruiting from three social media platforms: Instagram, Facebook, and Reddit. The second round took place shortly after, between April 9th, 2022 and April 16th, 2022, and focused on distributing the survey on a variety of online academic survey exchange groups.

All survey participants were screened to confirm that they were older than 18 prior to completing the survey.

The survey consisted of a total of 21 questions distributed amongst five sections, plus additional space for participants to ask questions or request to be contacted (**Supplemental; Online Privacy Survey**). The sections were as follows:

1. **Informed Consent** (*Q1*): Participants were provided with information about the study, the study's privacy and confidentiality practices, and contact information. This consent form had to be digitally signed before completing the survey.
2. **Demographic Information** (*Q2 – Q7*): Participants were asked to provide basic information such as their gender, age, field of study/work. All questions were optional.
3. **Knowledge Basis** (*Q8 – Q15*): Participants completed a short quiz to determine what they already knew about online privacy and tracking mechanisms.
4. **Online Usage/Habits** (*Q16 – Q18*): Participants were asked to provide the details of their daily internet usage.
5. **Privacy Perspectives** (*Q19 – Q21*): Participants were asked how comfortable they were with specific data being collected online and how necessary they perceived this collection to be. They then ranked their agreement with a series of privacy-focused statements.

All data was synthesized without explicit personal identifiers or unique ideas. Similarly, all data is presented in an aggregated manner.

In Person Interviews

Six survey respondents were also selected to complete in-person interviews. All interviewees were asked to expand on their survey responses, particularly focusing on the following prompts:

1. Online activity and habits
2. Concerns about and perceptions of online privacy
3. Knowledge of cookies and whether they enable them
4. Knowledge and perceptions of browser fingerprinting
5. Weighing the benefits and risks of browser fingerprinting.

Interested interviewees were also directed through a hands-on experience of browser fingerprinting. After viewing whether their browser fingerprint data was unique via a preexisting online resource (amiunique.org), participants were then asked to complete an activity where they requested and analyzed their Facebook data, particularly that surrounding the advertisements and keywords that Facebook's algorithm had determined were of interest. To conclude the user study, interviewees were given the opportunity to browse the internet as normal, during which time a proxy was used to intercept web traffic. This allowed them to then view the data that had been collected on them by the different sites, a browser fingerprinting experience that allowed interviewees to be both the trackers and the tracked.

Results

The Participants

Responses were collected from a total of 67 participants, with 42 recruited during round 1 of survey distribution and the remaining 25 recruited during round 2. One response was discarded due to the participant failing to properly fill out the consent paperwork.

Of the participants, 63% – the majority – identify as women, while 28% identify as men, and 6% identify as non-binary. Most were young adults between the ages of 18 and 23, while 35% were between 23 and 50 and the remaining 8% were older than 50 (**Supplemental; Table S1**). Participants represent a total of 17 countries, with those from the U.S. largely coming from New York, although 14 other states were represented as well (**Fig. 3**). Participant's backgrounds varied, as less than a quarter indicated an engineering or technical field of study or work, with answers ranging from the natural sciences and business to history and art (**Supplemental; Table S2**).

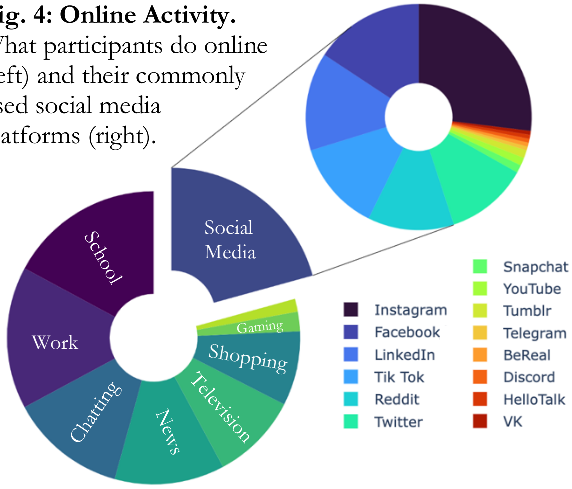


Fig. 4: Online Activity. What participants do online (left) and their commonly used social media platforms (right).

Participants tend to spend between 4 and 10 hours online per day and the most common online activities include social media, school, work, talking with others, and reading the news. In terms of social media, the most commonly used social media platforms among the participants are Instagram, Facebook, LinkedIn, Tik Tok, Reddit, and Twitter.

Knowledge Basis

All but one participant were familiar with browser cookies as a tracking technique, while significantly fewer indicated awareness of browser fingerprinting (30%) or supercookies (14%). However, in a multiple-choice question

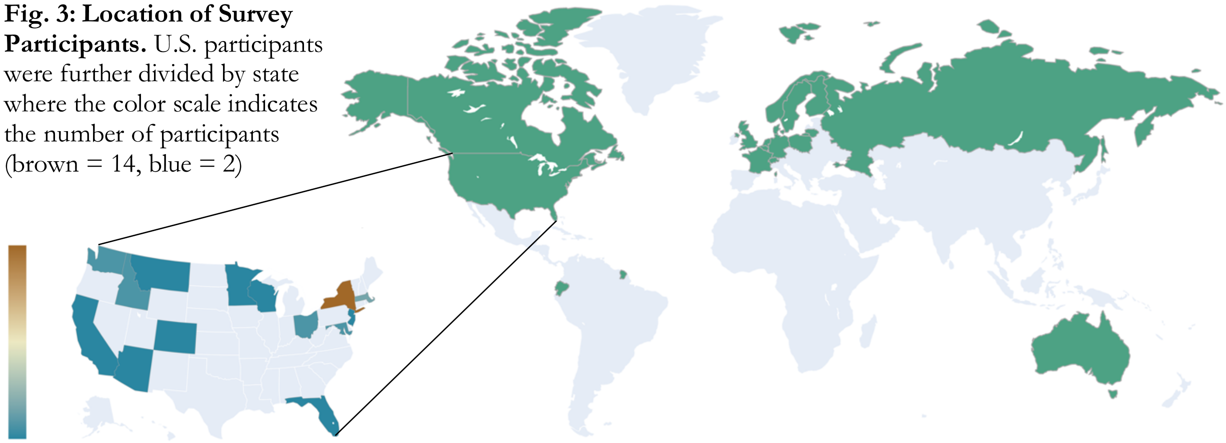


Fig. 3: Location of Survey Participants. U.S. participants were further divided by state where the color scale indicates the number of participants (brown = 14, blue = 2)

regarding the original intent of cookies, the vast majority of participants answered incorrectly that cookies were either intended for targeted advertisements or for tracking users. Only seven of the participants provided the correct answer referencing virtual shopping carts.

Not only could they identify browser cookies from a list, but participants were largely able to provide an adequate definition. While some definitions focused on *what* browser cookies were as "a site's memory", others focused on the *use* or *impact* of cookies. Common mentions were the creation of a "harmonious internet experience", their contribution to "more personalized ads", and their use to "track [user's] data and sell it to other companies". On the other hand, though, participants were largely unable to define browser fingerprinting. The vast majority wrote that they "don't know" or had "never heard of this", while others admitted to guessing by "using context clues".

Focusing in on browser fingerprinting, 65% of participants believed they could protect themselves from fingerprinting by blocking web trackers, turning off cookies, and/or using incognito mode, while an additional 20% indicated that at least one of these methods would work. Only 15% correctly realized that none of these techniques were sufficient to prevent oneself from being fingerprinted.

In the final portion of this knowledge basis "quiz", 52% of the participants indicated that they believed that U.S. federal law regulates data privacy. When then asked to select all the states they believed to have comprehensive data privacy laws, answers varied. California, New York, Washington, and Maryland were the most common selections (**Fig. 5**). In reality, though, there were only three correct answers: California, Colorado, and Virginia (*yellow outlines,* **Fig. 5**).

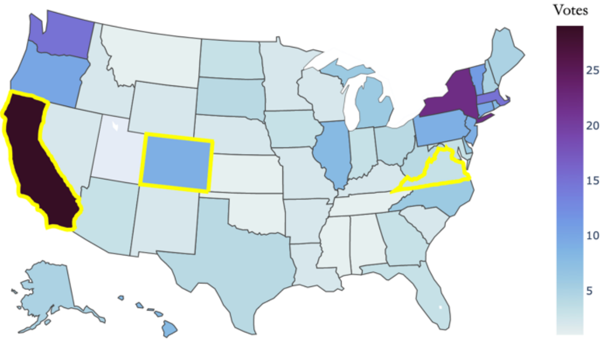


Fig. 5: Participant Knowledge of Privacy Laws. Participants selected the states they believed had comprehensive digital privacy laws. Choropleth map displays how many votes each state received while yellow outlines indicate the correct responses.

Privacy Perspectives

The list of data collected online, as presented to participants, was categorized into five tiers based on how "personal" the data was. In other words, the types of data were sorted based on each characteristic's ability to narrow down a population towards identifying a single individual. Tier 1 was defined as the least personal, meaning unlikely to be able to identify an individual, while Tier 5 was the most personal, containing data that could identify a single individual. The sorting of the tiers is as follows:

1. **Tier 1:** Screen Size, Device Model, Operating System, Browser Name, Installed Fonts, Time Zone
2. **Tier 2:** Browser Extensions, Installed Drivers
3. **Tier 3:** Gender, School/Job
4. **Tier 4:** Location, IP Address
5. **Tier 5:** Name, Email

Overall, a clear trend can be observed between comfort level and how personal the data is. As the information becomes more personal, comfort level decreases while discomfort level subsequently increases ($p \leq 0.01$; **Supplemental; Table S3** and **Figure S1**). Interestingly, though, there is no statistically significant trend between how necessary participant's perceived collection of each data type to be and its tier. Rather, necessity was just generally perceived to be low, with all data declared to be necessary only sometimes to seldom (**Fig. 6**).

Ultimately, most participants believe that they are unable to protect their privacy online and that their state/local laws also do little to

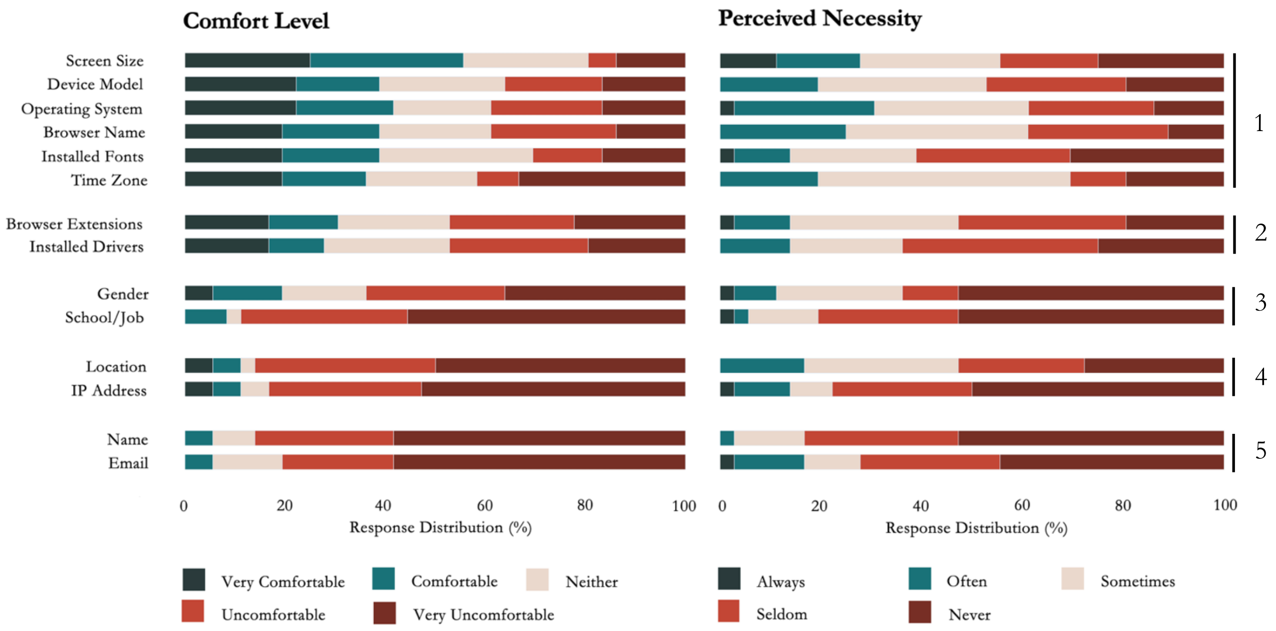


Fig. 6: Comfort with and Perceived Necessity of Data Collection. Participants ranked each form of data on a scale from very comfortable to very uncomfortable with its collection and on a scale from believing collection is always necessary to believing collection is never necessary.

adequately protect them either. If follows that most believe a browser fingerprint would be able to uniquely identify them. Most interestingly – and arguably most crucial – is that the vast majority had never heard of browser fingerprinting or learned ways to protect their privacy online prior to taking this survey, indicating a lack of transparency and educational resources.

Interviews

Resulting insights from in-person interviews further supported those from the online survey. Ultimately, three main takeaways can be realized:

1. **Most do take steps to protect their privacy online.** Whether it's using in-private browsing, using a VPN, or ensuring social media accounts are private, all expressed some direct intention to preserve their online privacy. One interviewee expressed that they were "very into Duo" while another expressed that they always "tried to manage [their] cookies."
2. **Despite this, most feel they are inadequately equipped to protect their privacy.** One participant expressed that "privacy…means do your best, but [they] don't really understand it" while others expressed that they "feel like [they] can't control it". There appears to be a clear knowledge gap between those who would consider themselves technological and those who labeled themselves as definitely "not a computer scientist"
3. **Their browser fingerprint is effective and experiencing this served as an important educational step.** All interviewees felt strongly that Facebook's description of the topics it used for their targeted advertisements reflected their personality. Many expressed how "it works, it works really well, because [they] often fall for the ads." After taking the survey, being interviewed, and completing the browser fingerprinting simulation, it was observed that participants could

more readily define both browser cookies and browser fingerprinting, indicating that experiences such as these are important educational steps towards learning to protect one's online privacy.

Discussion and Conclusion

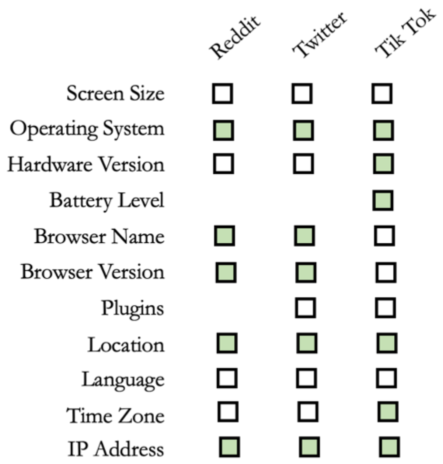
Through online user surveys and in-person interviews, regular internet users were recruited to provide insights into the implications of user perspectives and awareness of online tracking techniques. Ultimately, while there is a general understanding and concern that such techniques do exist, it appears that little is generally known about what cookies or browser fingerprints are truly doing or how one can protect oneself from being tracked.

A potential limitation of these study results is the relatively small sample size (n=66) – a product of the short time frame – and the online recruitment procedure with limited screening. However, similar studies have shown that online survey communities and services do provide generalizable results (Redmiles et. al., 2019). Survey data is further strengthened by the fact that both rounds of surveying, despite different time frames and strategies for recruitment, produce similar if not identical trends in data.

Ultimately, the insights gained through this study can be applied to direct the formation of education resources and advise future direction for privacy legislation. In taking the first step, outtakes of interviews have been compiled into a short documentary highlighting important contributions from interviewees to the discussion of online tracking and its implications for privacy. Moving towards the second application, results of this study can be synthesized in a few final thoughts.

They may be sweet but that doesn't hide the sour

While cookies and browser fingerprinting are inherently opportunities for privacy violations, they also provide a myriad of beneficial applications. When asked about the definition or intention of these technologies, results showed that these words generally have negative connotations for users, and as such, their responses and thoughts generally focus on the harmful aspects of these technologies. However, they can also be used for online activities that are commonly taken for granted, whether that's maintaining one's shopping cart, keeping one logged into their account, or even strengthening web security through fraud prevention on banking sites! Moving to the future, we must find ways to balance these benefits with the privacy risks, something that will likely come from a combination of user education, company transparency, and legislation development.



	Reddit	Twitter	Tik Tok
Screen Size	□	□	□
Operating System	🟩	🟩	🟩
Hardware Version	□	□	🟩
Battery Level			🟩
Browser Name	🟩	🟩	□
Browser Version	🟩	🟩	□
Plugins		□	□
Location	🟩	🟩	🟩
Language	□	□	□
Time Zone	□	□	🟩
IP Address	🟩	🟩	🟩

Fig. 7: Information Collected and Declared. A black box indicates that evidence of this data being collected was found. Green highlighting marks the data collection that is explicitly declared in the company's privacy policy.

Users are not fortune cookies

The average user does not go out of their way to investigate every website's policy to determine exactly what information is being collected. Instead, they leave their privacy in the hands of such websites and, as such, cannot predict what will happen with their data. Following up on reported social media usage, the privacy policies of Reddit, Twitter, and Tik Tok were compared to evidence of cookies and browser fingerprinting as found through the use of a proxy server (**Fig. 7**). In general, much information is left out of privacy policies, largely due to the use of vague language such as "Device Information" or "General Settings." Even if users were to investigate what data is being collected, the answer would not be satisfactory nor useful. Thus, it is recommended that websites and companies revisit their policies, clarify their language, and ensure that transparency is at the forefront of every discussion of user privacy.

Lawmakers need to be tough (on) cookies

Results indicating that most internet users' knowledge of the data privacy laws protecting them is little to none were surprising given that such laws do exist. However, this outcome is likely a result of their limitation and ineffectiveness. In the U.S., there is no single, overarching federal law that regulates data privacy. This has resulted in a complex entanglement of many smaller laws and a multitude of state laws. The strongest of these laws is California's Consumer Privacy Act, or CCPA, which mandates that websites must disclose their collection of third-party cookies. Additionally, websites must include a means through which users can explicitly opt out of this data collection (Bonta, 2018). However, the CCPA does have several shortcomings. Websites are not required to display a cookie banner, notice, or pop-up, making it the responsibility of everyday users to know that their data is being collected and to take the time to figure out how to disable cookies if they wish. The opt-out policy is also a major shortcoming, as most people will not take the time to do so.

What's most prominently missing, though, is regulation explicitly mentioning browser fingerprinting. Why require websites to declare their use of cookies and not browser fingerprinting when the two are used for – and can both successfully accomplish – the same online tracking? Not only should future legislation focus on strengthening regulations around cookies, but it should incorporate browser fingerprinting in a clear manner.

Ultimately, as far as company transparency and better legislation will go, user awareness will be irreplaceable. It appears that a main hindrance to better privacy online is a lack of knowledge about how one's data is being used and how one can control this. After seeing browser fingerprinting in action and exploring protective techniques, users responded with an increased sense of how their data was used and how they could better protect their own privacy online. Thus, this project accomplished its goal for these participants, and by compiling the interviews into a miniature documentary accompanied by a set of resources, it hopes to do the same for others as well.

References

Bhagyashree, R. (2019). All about Browser Fingerprinting, the privacy nightmare that keeps web developers awake at night. Packt>. Retrieved from https://hub.packtpub.com/all-about-browser-fingerprinting- the-privacy-nightmare-that-keeps-web-developers-awake-at-night/

Bonta, R. (2018). California Consumer Privacy Act (CCPA). State of California Department of Justice. Retrieved from https://www.oag.ca.gov/privacy/ccpa

Burgess, M. (2022). The Quiet Way Advertisers are Tracking Your Browsing. Wired. Retrieved from https://www.wired.com/story/browser-fingerprinting-tracking-explained/

Chen, B.X. (2019). 'Fingerprinting' to Track Us Online Is on the Rise. Here's What to Do. The New York Times. Retrieved from https://www.nytimes.com/2019/07/03/technology/personaltech/fingerprinting-track- devices-what-to-do.html

Conger, K., & Chen, B.X. (2022). A Change by Apple Is Tormenting Internet Companies, Especially Meta. The New York Times. Retrieved from https://www.nytimes.com/2022/02/03/technology/apple- privacy-changes-meta.html

Federal Communications Commission. (2016, March 7). FCC Settles Verizon "Supercookie" Probe. Retrieved from https://www.fcc.gov/document/fcc-settles-verizon-supercookie-probe

Jackson, T. (1996). This bug in your PC is a smart cookie. The Financial Times. Screenshot Retrieved from https://baekdal.com/thoughts/the-original-cookie-specification-from-1997-was-gdpr-compliant/

Kihn, M. (2018). Cookies, Chaos and the Browser: Meet Lou Montulli. Gartner. Retrieved from https://blogs.gartner.com/martin-kihn/cookies-chaos-and-the-browser-meet-lou-montulli/

Kristol, D. (1997). RFC2109: HTTP State Management Mechanism. Internet Engineering Task Force. Retrieved from https://datatracker.ietf.org/doc/html/rfc2109#section-8.3

Kristol, D. (2000). RFC2965: HTTP State Management Mechanism. Internet Engineering Task Force. Retrieved from https://datatracker.ietf.org/doc/html/rfc2965

Montulli, L. (1998). US5774670A: Persistent client state in a hypertext transfer protocol based client- server system.

Mowery, K. and Shacham, H. (2012). Pixel Perfect: Fingerprinting Canvas in HTML5. Department of Computer Science and Engineering, University of California, San Diego. Retrieved from https://hovav.net/ucsd/dist/canvas.pdf

Mozilla. (2022). Using HTTP Cookies. MDN Web Docs. Retrieved from https://developer.mozilla.org/en-US/docs/Web/HTTP/Cookies

Phillips, G. (2019, August 22). What Are Supercookies? Here's How to Remove Them Properly. Make Use Of. https://www.makeuseof.com/tag/what-are-supercookies-and-why-are-they-dangerous/

Redmiles, E.M., Kross, S., & Mazurek, M.L. (2019). How Well Do My Results Generalize? Comparing Security and Privacy Survey Results from MTurk, Web, and Telephone Samples. University of Maryland. Retrieved from https://drum.lib.umd.edu/bitstream/handle/1903/19164/CS-TR-5054.pdf?sequence=1

Roesner, F., Kohno, T., & Wetherall, D. (2012). Detecting and Defending Against Third-Party Tracking on the Web. USENIX Symposium on Networked Systems Design and Implementation. Retrieved from https://www.usenix.org/system/files/conference/nsdi12/nsdi12-final17.pdf

Singer, N. and Chen, B. X. (2015, January 25). Verizon's Mobile 'Supercookies' Seen as Threat to Privacy. The New York Times. https://www.nytimes.com/2015/01/26/technology/verizons- mobile-supercookies-seen- as-threat-to-privacy.html

Soltani, A., Canty, S., Mayo, Q., Thomas, L., and Hoofnagle, C.J. (2009). Flash Cookies and Privacy. School of Information, University of California, Berkeley. Retrieved from https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862

Wickramasinghe, S. (2021). A Complete Guide to Browser Fingerprinting – What it is and How it Affects You. Privacy Affairs. Retrieved from https://www.privacyaffairs.com/browser-fingerprinting/

Supplemental Information

Table S1: Gender and age of survey participants. Demographics are reported both as an absolute count and as a percentage of the total participants. Note that participants could select multiple genders, so the absolute counts may not sum to the total number of participants.

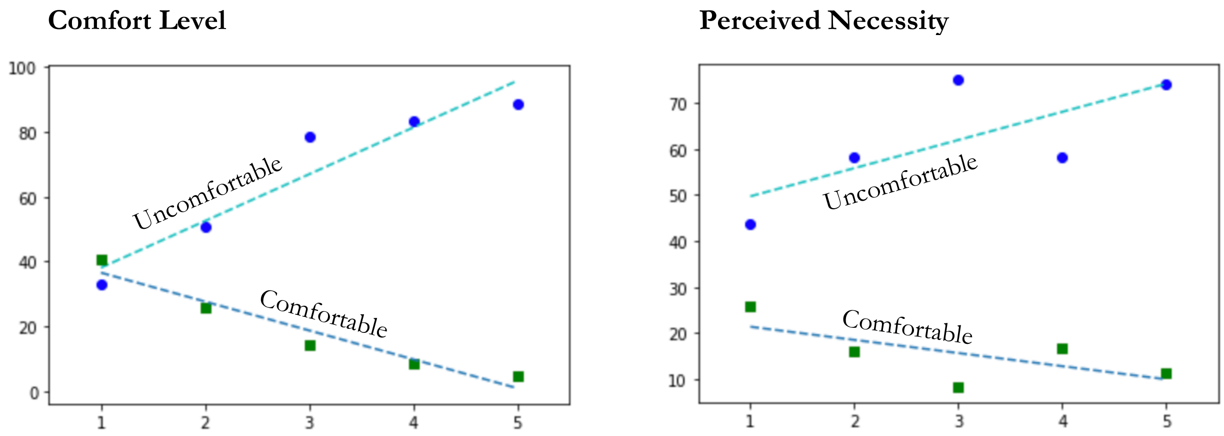
		Survey Responses (Total = 66)	
		Count	**Percentage (%)**
Gender	Woman	42	63
	Man	19	28
	Non-Binary	4	6
	Prefer not to Answer	2	3
Age	18 – 23	38	57
	23 – 30	17	26
	30 – 40	6	9
	40 – 50	0	0
	50 – 60	4	6
	60 +	1	2

Table S2: Participant Fields. Individual answers were manually categorized. Number of responses for each are reported both as an absolute count and as a percentage of total participants.

Category of Work/Study	**Count**	**Percentage (%)**
Engineering / Comp. Sci.	15	23
Natural Sciences	10	15
Language / Anthropology	7	11
History / Politics	3	4
Art / Design / Writing	6	9
Business	9	14
Math / Economics	3	4
Education	5	8
Retired	4	6
Prefer not to Answer	4	6

Table S3: Comfort and Perceived Necessity of Data Collection. Specific forms of data have been aggregated into Personal Information Tiers, with Tier 1 being the least personal data and Tier 5 being the most personal data. Regression analyses of trends in each variable with increasing tier are also presented.

Personal Information Tier	Comfortable (% Responses)	Uncomfortable (% Responses)	Always/Often Necessary (% Responses)	Never/Seldom Necessary (% Responses)
1	40	33	26	44
2	26	51	16	58
3	14	79	8	75
4	8	83	17	58
5	5	89	11	74
Regression Analyses				
Slope	-8.91	14.41	-2.85	6.11
Intercept	45.43	23.61	24.22	43.59
Multiple R	-0.97	0.95	-0.67	0.74
R Square	0.94	0.90	0.45	0.55
Std. Error	1.32	2.75	1.82	3.20
P-value	< 0.01	0.01	>> 0.01	>> 0.01



Comfort Level

Perceived Necessity

Figure S1: Linear Regression Between Comfort Level / Perceived Necessity and Personal Information Tier. While there is a significant positive correlation between discomfort and how personal the data is and a significant negative correlation between comfort and how personal the data is, there are no significant trends between perceived necessity and personal information tier.

ONLINE PRIVACY SURVEY

Consent to Participate in Research Study

I. Why is this study being done?

We are doing this research study to better understand how people think about and/or protect their privacy online. We hope to learn more about general awareness and opinions of tracking techniques, with a specific focus on browser fingerprinting.

II. What will I be asked to do if I choose to be in this study?

We will ask you to either complete a survey or answer questions verbally.

III. What about my privacy?

Every effort will be made to keep your personal information confidential. However, we cannot guarantee total privacy.
Survey data collected will be given a code number and separated from your name or any other information that could identify you, if provided. Only the Principal Investigator and the study staff will be able to see this information.
The research information that is shared with people outside of the study team will not include your address, telephone number or any other direct identifier besides your name unless disclosure of the information is required by law or you have authorized the disclosure.
The following people and/or agencies will be able to look at, copy, use and share your research information:

- The investigator and Barnard College study staff and other professionals who may be evaluating the study.
- Authorities from Barnard College and Columbia University, including the Institutional Review Board ('IRB'). An IRB is a committee organized to protect the rights and welfare of people involved in research.

IV. Are there any risks?

There are no physical risks related to participating in this research study.
You may choose to skip questions if they make you uncomfortable.
A risk of taking part in this study is the possibility of a loss of confidentiality or privacy. Loss of privacy means having your personal information shared with someone who is not on the study team and was not supposed to see or know about your information. The study team plans to protect your privacy.

V. Are there any benefits?

You may or may not receive personal benefit from taking part in this study. The possible benefits of taking part in this study include:

- Learning more about browser fingerprinting and online privacy in general
- Acquiring strategies to better protect your privacy online

By participating in this study, you will be contributing to educational resources for others to potentially have the same benefits as well.

Q1: Do you consent to participate in this research study?

○ Yes
○ No

Demographic Information

All of the following questions are optional. However, they provide very useful information for us, so please answer as many as you feel comfortable. If you choose not to, please select the "Prefer not to Answer" option.

Q2: Gender Identity (Select all that apply)

□ Woman
□ Man
□ Non-Binary
□ Other
□ Prefer not to Answer

Q3: Age

○ 18 – 23
○ 23 – 30
○ 30 – 40
○ 40 – 50
○ 50 – 60
○ 60 +
○ Prefer not to Answer

Q4: Are you a student?

○ Yes
○ No
○ Prefer not to Answer

Q5: If a student, what is your field of study?

____

Q6: If not a student, what is your career field?

____

Q7: State you live in. If you are outside of the U.S., you can list your province or country.

____

Knowledge Basis

The following questions will quiz you on your current knowledge about online privacy. PLEASE DO NOT GOOGLE ANY OF THE QUESTIONS. Simply answer them to the best of your ability so that we (the researchers) can understand what prior knowledge you may have.

Q8: Which of the following are online tracking techniques (select all that apply)

□ Cookies
□ Supercookies
□ Megacookies
□ Footprinting
□ Fingerprinting
□ Other: _____

Q9: Define Browser Cookies

____

Q10: Define Browser Fingerprinting

____

Q11: Cookies were Originally Developed to…

○ Track visitors to a website
○ Allow advertisements to be targeted to user interests
○ Implement virtual shopping carts

Q12: Which of the following can any website collect about us? (Select all that apply)

□ Screen Size
□ Operating System
□ Browser Name
□ Browser Version
□ Browser Language
□ Time Zone
□ Battery Level
□ Browser Extensions Installed
□ Font Types Installed
□ Device Memory

Q13: Which of the following can protect you from browser fingerprinting?

□ Turning off/deleting cookies
□ Blocking web trackers
□ Using Incognito mode
□ All of the above
□ None of the above

Q14: In the U.S., federal law regulates data privacy

○ True
○ False

Q15: Select all of the states that have comprehensive data privacy laws

[list of the 50 U.S. States]

Online Usage and Habits

Q16: How many hours would you estimate that you spend online per day?

○ 0
○ 1
○ 2 – 3
○ 4 – 6
○ 6 – 10
○ > 10

Q17: Which of the following activities do you generally do during your time online?

□ School
□ Work
□ Social Media
□ Television
□ News
□ Talking with Others
□ Shopping
□ Other: _____

Q18: Which of the following social media platforms are you regularly active on?

□ Facebook/Meta
□ Instagram
□ Twitter
□ Tik Tok
□ Reddit
□ LinkedIn
□ Other: _____

Privacy Perspectives/Concerns

Q19: How comfortable are you with websites collecting each of the following pieces of information?

	Very Comfortable	Comfortable	Neither	Uncomfortable	Very Uncomfortable
Screen Size	□	□	□	□	□
Device Model	□	□	□	□	□
Operating System	□	□	□	□	□
Browser Name	□	□	□	□	□
Browser Extensions	□	□	□	□	□
Installed Fonts	□	□	□	□	□
Installed Drivers	□	□	□	□	□
Time Zone	□	□	□	□	□
Location	□	□	□	□	□
IP Address	□	□	□	□	□
Name	□	□	□	□	□
Email	□	□	□	□	□
Gender	□	□	□	□	□
School/Job	□	□	□	□	□
Birthday	□	□	□	□	□

Q20: How necessary do you think it would be for websites to collect each of the following pieces of information?

	Always	Often	Sometimes	Seldom	Never
Screen Size	□	□	□	□	□
Device Model	□	□	□	□	□
Operating System	□	□	□	□	□
Browser Name	□	□	□	□	□
Browser Extensions	□	□	□	□	□
Installed Fonts	□	□	□	□	□
Installed Drivers	□	□	□	□	□
Time Zone	□	□	□	□	□
Location	□	□	□	□	□
IP Address	□	□	□	□	□
Name	□	□	□	□	□
Email	□	□	□	□	□
Gender	□	□	□	□	□
School/Job	□	□	□	□	□
Birthday	□	□	□	□	□

Q21: To what extent do you agree with the following statements?

	Strongly Agree	Agree	Neutral	Disagree	Strongly Disagree
I can protect my privacy online	□	□	□	□	□
I actively work to protect my privacy online	□	□	□	□	□
I usually allow cookies when a website asks	□	□	□	□	□
I read the information to check what information each website is collecting about me	□	□	□	□	□
Using Incognito Mode (or private browsing) can protect me from browser fingerprinting	□	□	□	□	□
I often use incognito mode (or private browsing)	□	□	□	□	□
Using a VPN (Virtual Private Network) can protect me from browser fingerprinting	□	□	□	□	□
I often use a VPN	□	□	□	□	□
A browser fingerprint could uniquely identify me	□	□	□	□	□
My state/local laws protect my privacy online	□	□	□	□	□
I have previously received information about browser fingerprinting and how to better protect myself	□	□	□	□	□

Final Question

Q22: Do you have any additional comments, questions, or concerns?

____

Q23: If you would like a researcher to follow up with you, please enter you remail address below.

____