Rethinking Fingerprinting: An Assessment of Behavior-based Methods at Scale and Implications for Web Tracking

Kyle Crichton
Georgetown University
kyle.crichton@georgetown.edu

Lorrie Faith Cranor
Carnegie Mellon University
lorrie@cmu.edu

Nicolas Christin
Carnegie Mellon University
nicolasc@cmu.edu

Abstract

Most common forms of web tracking fail to maintain the continuity of a user's identity over long periods of time: cookies get deleted, IP addresses are reassigned, attributes used for browser fingerprinting change. These *identity discontinuities* help prevent adversaries from conducting persistent long-term tracking. In fact, many privacy-enhancing technologies (e.g., automatic cookie deletion, use of proxy servers, fingerprint obfuscation) are predicated on the ability of identity discontinuities to disrupt an adversary's tracking capability. While only evaluated on a limited scale, behavioral fingerprinting—identifying users based on habitual patterns in their web browsing—may provide adversaries the key to linking users' identities across these discontinuities.

To assess this potential threat, we provide an analysis of behavioral fingerprinting at scale, with over 150,000 users across two years, and the first assessment of the impact of these techniques on user anonymity online. Overall, we find that behavioral fingerprints are relatively unique, with most browsing sessions retaining little to no anonymity even at scale. Furthermore, users' behavioral fingerprints are consistent, evolving slowly over the course of months to years. Together, these findings satisfy the preconditions for effective identity linking. We go on to demonstrate that optimal performance is achieved when an adversary can observe 15–25 browsing sessions prior to a discontinuity and 10–15 sessions after. However, an adversary can eliminate 84–95% of a user's anonymity having observed just a single session pre- and post-discontinuity. After a discontinuity occurs, a user loses an average of 78–85% of their anonymity within the first 60 seconds of browsing and 90% of their anonymity within the first 10 minutes—largely negating the anonymity gains of privacy protections that induce discontinuities. We find that visiting fewer web pages, diversifying the websites visited, and avoiding niche content can help a user's browsing remain anonymous. Finally, we demonstrate that the combination of behavioral and browser fingerprinting can outperform each method individually, achieving an F1 score of 0.869 across 100,000 users.

Keywords

web tracking, fingerprinting, privacy, online behavior

1 Introduction

Many privacy-enhancing technologies designed to prevent web tracking are predicated on the ability to disrupt the continuity of an adversary's tracking capability. Indeed, even when these tools are not actively used, common web tracking techniques are limited in their ability to link the identity of a given user as the identifiers used to track them (e.g., browser cookie, IP address, browser fingerprint) change over time, events we will refer to as *identity discontinuities*. While these methods remain effective at tracking users' behavior between discontinuities—potentially for days, weeks, or months at a time—this challenge limits the ability of an adversary to carry out persistent long-term web tracking across these time periods. However, we find that an often overlooked technique based on the patterns in a user's web browsing, a user's *behavioral fingerprint*, may provide trackers with a privacy-invasive solution.

An identity discontinuity occurs when a change is made to the original identifier used for tracking such that an adversary can no longer maintain the connection between the user's identity before and after the change occurs. As a result, the adversary retains two disconnected sets of browsing data, each associated with a different identifier, that in reality both belong to the same user. A discontinuity can result from a wide variety of events: cookies can be cleared from the browser or are deleted upon reaching their expiration; the user's router can be assigned a new IP address or the user may connect to a different network entirely; and browser fingerprints can change due to software updates being applied, system settings being modified, or simply connecting to an external monitor. Not only do these events occur naturally over time, many privacy-enhancing technologies are designed to introduce discontinuities to disrupt an adversary's tracking—including intelligent tracking protection that dynamically clear cookies [112], proxy servers used to mask IP addresses, and defenses that selectively or randomly modify browser fingerprint attributes [12, 37, 38, 51, 59, 81, 100, 101].

In this paper, we examine how behavioral fingerprinting can be applied by an adversary to overcome the identity discontinuity problem, potentially subverting protections based on it. Behavioral fingerprinting is based on individuals' browsing habits being relatively unique, but also habitual [28, 31, 83]. Initial evidence suggests that the consistency and uniqueness of online behavioral fingerprints, defined by the websites we visit, can be used for identification purposes just like our physical fingerprints [9, 14, 43, 46, 54, 84, 85, 109, 118]. Yet, behavioral fingerprinting is often overlooked by privacy researchers because, unlike other methods of web tracking, it cannot instantly identify a user and instead requires observing their behavior over a period of time. However, where the identifiers used in other tracking techniques change when an identity discontinuity occurs, behavioral fingerprints remain consistent. Therefore, an adversary—who is already collecting behavioral fingerprinting information as an inherent part of web tracking—can take advantage of this consistency to link disconnected identifiers belonging to the same user. Such a capability would provide an adversary

Proceedings on Privacy Enhancing Technologies YYYY(X), 1–17

https://doi.org/XXXXXXX.XXXXXXX

with a more comprehensive picture of a user's browsing history over time, raising the specter of evercookies [92] and enabling more persistent and privacy-invasive long-term web tracking.

Our work examines the effectiveness of linking user identities via behavioral fingerprinting using a browsing dataset collected from 150,000 average daily users over the course of two years. In comparison, previous evaluations were limited to samples of several thousand users observed over a period of a few weeks. To the best of our knowledge, we present the first assessment of the impact of this technique on anonymity and the combination of behavioral and browser fingerprinting for identity linking. Using our unique dataset, and validating our results using data from two previous studies [46, 109], we examine the following research questions:

- How unique are behavioral fingerprints between users?
- How consistent are behavioral fingerprints over time?
- How does the amount of information an adversary observes before and after a discontinuity affect performance?
- How quickly do users lose anonymity after a discontinuity?
- Can combining behavioral and browser fingerprinting enhance an adversary's ability to link user identities?

In our analysis, we find that behavioral fingerprints satisfy the preconditions for effective identity linking in that they are relatively unique and consistent. Linking performance declines as the number of users increases, but the marginal performance loss becomes minimal beyond 25,000 users. Worse, most browsing sessions retain little to no anonymity even at scale. While recent knowledge of a user's behavioral fingerprint is valuable, we find that fingerprints evolve slowly over the course of months to years. Optimal performance is obtained when an adversary observes a user's browsing for 15–25 days prior to a discontinuity occurring, however these methods can still reduce anonymity by at least 84–95% having only observed one previous browsing session. Similarly, being able to observe a user for more than 10–15 days after a discontinuity provides the best results, enabling the model to identify a given user in 71.5% of cases and reduce anonymity by 99.9% for 87.9% of sessions. Post-discontinuity, we find that users on average lose 78–85% of their anonymity within the first 60 seconds of browsing and 90% anonymity within the first 10 minutes. Visiting a greater number of web pages but across a smaller number of websites, particularly those that are more niche and that you tend to visit regularly, increases the likelihood that anonymity is lost while browsing. Combining our results with that of a recent study on browser fingerprinting [5] we show that a combination of fingerprinting techniques outperforms each linking method individually, achieving a theoretical F1 score of 0.869 across 100,000 users.

2 Related Work

In the following sections we cover related work pertaining web tracking at large, followed by tracking techniques based on cookies, IP addresses, browser fingerprints, and behavioral fingerprints.

2.1 Web Tracking and Privacy

Over the past two decades, exponential growth in the number of trackers, the third parties they represent, their coverage, and the diversity of their techniques [62] has led to a near-ubiquitous state of tracking online [64, 88, 91, 106]. While most data collection is centralized among a handful of companies [17, 35, 48, 62, 64, 91], data is increasingly aggregated and shared among a large number of interconnected third-parties [11, 49, 93]. The centralization of tracking coverage has grown over time [55, 62, 93], driven by major corporate acquisitions in recent years [13]. These trends have been consistent across personal computers and mobile devices [23, 119].

Tracking user behavior across the internet, and the data collected as a result, is primarily employed for use in Online Behavioral Advertising (OBA), an estimated $566B industry in 2022 [95]. Information about web page visits enables advertising networks and data aggregators to draw inferences about individuals and their interests in order to serve them more relevant advertisements. Recent measurements have found that 63–78% of advertisements served are targeted to users based on these methods [10, 22, 65].

Users have specifically expressed concerns over the sensitivity of data collected [24, 61, 71]. Prior research has found that tracking on sensitive websites, while less pervasive than on mainstream sites, is still highly prevalent [50, 91]. This includes tracking on websites relate to healthcare [50, 80, 91, 114, 120], mental health [91], abortion [40], addiction [91], e-governance [42, 90], political affiliation [50], pornography [66, 108], religion [50], ethnicity [50], sexual orientation [50, 91], and gender identity [91]. Additional research has pointed to issues where web tracking can lead to algorithmic bias [4, 6, 7, 30, 57, 97, 99, 116], price discrimination [45, 72], and chilling effects on internet use and free speech [70, 96].

2.2 Cookie-based Tracking

Browser cookies have long been the predominant form of web tracking. Cookies are a form of stateful tracking in which the user is assigned a unique identifier by the tracker, that identifier is stored in the user's browser and then is passed back to the tracker when visiting a site where their trackers reside. First-party cookies belong to the website a user visits, while third-party cookies belong to external entities that have embedded content within the first-party website [34]. Embedded trackers that can spawn third-party cookies include web beacons [67, 89], and embedded scripts [52]. Large-scale analysis has found that 90% of websites leak some form of user data to external entities, 80% load JavaScript from external parties, and 60% spawn third-party cookies [64]. In addition, identifiers are often exchanged between multiple third-parties in a process called "cookie syncing," extending the reach of many trackers [1, 34].

Although cookie-based tracking is near-ubiquitous, cookies themselves do not last forever. Users can interrupt the continuity of cookie-based tracking by clearing the cookie cache in their web browser and can temporarily disable cookies by using private browsing mode. Furthermore, cookies are automatically deleted after reaching their expiration date. Previous work shows that 10% of cookies are reset daily, about 20% weekly, and just over 30% monthly [29]. These rates are likely higher for certain web browsers such as Brave, Firefox, and Safari who enforce tighter restrictions on cookie creation and expiration [27]. While the use of flash and other local storage objects to respawn deleted cookies—enabling persistent "evercookies"—represented a substantial threat to privacy 10 years ago [8, 69, 94], these methods have largely become obsolete as web browsers and related technology have evolved [92].

Protections against cookie-based tracking also exist. Chief among them are ad blockers, tools that prevent many of the third-party cookies and embedded trackers from being loaded when a user visits a site. However, the all or nothing approaches to blocking trackers employed by these tools are in direct tension with the expressed needs of users who desire a balance between the benefits of personalization and associated privacy harms that stem from web tracking [2, 24, 61, 70, 71, 105, 113]. Tools that dynamically restrict or clear cookies at specified intervals, allow for some selective and temporary tracking [112]. Also partially addressing this need is the increased use of cookie consent mechanisms on websites, driven in large part by legislation under GDPR [36] and CCPA [18]. However, studies of the prevalence [47] and implementation [16, 44, 107] of these mechanisms have shown that they largely fail to provide meaningful user consent. While cookies remain an effective and widely-used form of tracking, these trends have incentivized web trackers to increasingly turn to alternative, stateless tracking mechanisms such as browser fingerprinting [86].

2.3 IP-based Tracking

Tracking users based on their IP address represents an alternative to the use of cookies that leverages the IP address that is assigned to the user's home network by their Internet Service Provider (ISP). The user's IP address, which is publicly visible in IP packet headers, identifies the source of web requests and is used to route responses back to the correct destination. Tracking users via IP address can occur through passive network observation [25], malicious DNS resolvers [46], and embedded trackers such as web beacons [53].

The challenge with IP-based tracking is that IP addresses tend to change at relatively frequent rates. Although static IP addresses exist, primarily for large businesses, most IP addresses are dynamically assigned by ISPs using Dynamic Host Configuration Protocol (DHCP). Recent work shows that 90% of global IP addresses are retained for less than 10 days. Average retention in the United States is higher at 18.93 days [73].

Defenses against IP-based tracking include the use of virtual private networks (VPNs) and the Tor network. In both cases, a user's web traffic is encrypted and routed through one or more proxy servers before being passed along to the destination website. From the perspective of an embedded tracker or a DNS resolver the source of the web request appears to be the last proxy server the request is forwarded from. The primary difference between these protections is that VPNs typically rely on a single proxy server, whereas Tor uses onion routing, passing traffic through a series of proxies. The latter makes it more challenging for network observers, particularly ISPs, to trace observed requests back to its source.

2.4 Browser Fingerprinting

Unlike previous methods, device or browser fingerprinting is considered stateless as it does not rely on identifiers that are assigned to the user. These techniques piece together several bits of information about a user's device to reduce the pool of possible individuals and infer their identity. These attributes include information stored in the HTTP header [32], JavaScript and Flash configuration [32, 74, 77], the Canvas [1, 75] and WebGL [21, 75] APIs, OS-level information [3, 15], and device hardware [79].

Browser fingerprinting on top websites has grown from 0.4% of the top 10,000 sites in 2013 [82], to 2.6% of the top 100,000 sites in 2016 [34], to over 10.0% of the top 100,000 sites in 2021 [51]. Where cookie-based tracking can often be detected on a web page, passive fingerprinting techniques can be invisible to web measurements used by privacy researchers [68]. Furthermore, inferences based on data contained in HTTP header information, or through various side-channel attacks, can be employed by observers on a network [32]. As such, these findings may underestimate the prevalence of browser fingerprinting, a possibility supported by recent findings that 69% of the top 10,000 sites collect at least one attribute that can be used for browser fingerprinting [3].

Measurements of the effectiveness of browser fingerprinting have declined over time. Early studies found that combinations of device and browser attributes could be used to uniquely identify between 89.4% [60] and 94.2% [32] of desktop users. However, more recent work has found that only 35.7% of profiles based on these attributes are unique. This change has been attributed to an evolution in web technology, in particular the progressive removal of browser plugins that has made less information about a device's configuration visible [41]. Based on these findings, subsequent work has examined fingerprints that use a much larger set of attributes, upwards of a ten-fold increase over previous work. Previous work has found that these fingerprints are much more unique, rivaling that of initial measurements [5, 87]. However, with a greater number of attributes, these fingerprints are more likely to change over time. Examples include attributes pertaining to the device's screen, which changes when connected to an external monitor; the browser window, which varies as a user resizes the window; and audio output, which can change with the use of headphones or other external audio devices. Research has demonstrated that this leads individual's browser fingerprints to change frequently, within days or hours—sometimes within the same browsing session [5, 63, 111].

Unlike cookies or IP addresses, which change completely when an identity discontinuity occurs, usually only a subset of attributes that comprise a browser fingerprint changes [5]. Among small populations (1,000–2,000 users), these methods have been able to link a user's identity across discontinuities for an average of 51 days [111] and for smaller subsets (44.4% of the sample examined) up to 81 days at a time [87]. However, these methods have been shown to perform poorly at scale [63]. Despite these limitations, recent measurements found that 3.8% of the top 30,000 websites employ browser fingerprinting to respawn deleted cookies [39].

A notable case that we will return to later is an assessment of browser fingerprinting for use in authentication conducted by Andriamilanto et al. [5]. Using a 262-attribute fingerprint, the authors propose a method of linking that can distinguish between fingerprints from the same browser and those from different browsers with a 0.61% false positive and negative rate. At first glance this appears to solve the identity discontinuity problem. However, the authors undersample comparisons between fingerprints of different uses to establish a 1:1 ratio with comparisons between the same user. Since there are many orders of magnitude more comparisons between fingerprints of different users than there are between those of the same users, even though the false positive rate is low the method would generate many times more false matches than it

would true matches. In practice, such a system would be impractical. However, if the pool of potentially confounding fingerprints could be narrowed (i.e. the anonymity of a given user could be reduced) then the proposed system could be viable.

Several protections against browser fingerprinting have been proposed. One set of methods includes tools that selectively modify fingerprint attributes requested by websites to mask the user [12, 37, 38, 51, 59, 81, 100, 101]. However, in some cases the additional uniqueness created by these tools has been shown to help fingerprinters, rather than obfuscate users [110]. To combat this, researchers have proposed methods of introducing random noise rather than selective changes to fingerprint attributes [19, 58]. An alternative approach taken by some browsers [76, 98] and tools [20, 117] have been to present a homogeneous fingerprint, making the device indistinguishable from others.

2.5 Behavioral Fingerprinting

The focus of this study is on a related, but less studied, technique known as behavioral fingerprinting. While browser fingerprinting relies on attributes of the user's device for identification, behavioral fingerprinting infers a user's identity based on aspects of a user's online behavior [9, 14, 43, 46, 54, 56, 109, 118]. These methods have garnered less attention from researchers as they lack instantaneous identification, instead requiring observation of a user's activity over a period of time, typically 24 hours in previous work [9, 43, 46, 54, 109, 118], to match fingerprints. However, the advantage of behavioral fingerprints is their ability to provide long-term continuity in tracking. Where identifiers used in other tracking techniques (e.g., cookies, IP addresses, browser fingerprints) change over time, individuals' browsing habits remain relatively consistent [28, 31, 83]. As such, behavioral fingerprinting could serve as a complement to other tracking techniques, helping to link different identifiers belonging to the same user over time.

Conceptually, behavioral fingerprinting traces back to research examining the concept of an "average user," an idea that has persisted in software development and research design despite lacking in empirical evidence [33]. Several studies have found that the way in which an individual user browses the web is relatively unique in terms of navigational behavior [83], browsing history [14, 84, 85], and even within-session browsing behavior [31]. Taking these findings a step further, researchers have conducted several studies to measure the effectiveness of behavioral fingerprinting [9, 14, 43, 46, 54, 109, 118]. This work has typically examined linking users between different 24-hour [9, 43, 46, 54, 109] or 1-week [14, 109] periods, though Herrmann et al. evaluated periods ranging from 1 minute (34% accuracy) to 28 days (98% accuracy) [46]. In these studies, user activity is primarily represented as a frequency vector of the website domains visited by the user and are transformed using Term-Frequency Inverse Document Frequency (TF-IDF) to account for commonly-visited sites. These studies and their results are summarized in Table 1.

With overall accuracy ranging from 50% to 88.2%, these studies provide evidence that behavioral fingerprinting is viable. However, most of these studies have evaluated relatively small pools of users over short periods of time. The largest set of users evaluated was 19,263 users, but only over the course of 2 weeks [14]. The longest study duration was a year, but only evaluated data from 100 participants [118]. As such, we do not know how well these methods perform under realistic conditions where the number of users is very large and adversaries have months to years worth of prior data to rely on. Various studies have examined how the accuracy of re-identifying users changes based on the length of observation [46, 109, 118], amount of activity observed [14], number of users [14, 118], feature selection strategy employed [14, 46, 54, 118], and classification model used [46, 109]. Further insights provided include an analysis of the theoretical capabilities of various third-party entities based on their tracking coverage across top websites [14] and the effectiveness of possible countermeasures [46].

Building on these studies, our work presents several novel contributions. First, we leverage our larger dataset to evaluate the effect of users, time, and information on a scale that previous work could not. Second, we present insights into the effect of behavioral fingerprinting on anonymity, not just uniqueness or re-identification, and the factors driving anonymity loss. Third, we evaluate the potential combination of behavioral and browser fingerprinting.

3 Methods

In the following sections we cover our experimental design, datasets, data processing, classifier design, and evaluation metrics.

3.1 Experimental Design

In our experiments, we simulate the role of an adversary who, in the course of web tracking, attempts to link identifiers across an identity discontinuity using behavioral fingerprinting. In this scenario, the adversary observes a set of browsers in time period B whose identifiers they did not previously observe in time period A. This indicates that either the tracker has never observed this user before or an identity discontinuity has occurred resulting in a previously observed user being associated with a new identifier. We model all our experiments as a classification problem in which the adversary attempts to predict the identity of an unknown user whose browsing is observed in time period B, based on the set of previously observed users, and their browsing, in time period A. Inherently, this technique will be less effective in linking the identities of users who exhibit rapid, intense changes in browsing behavior and those who have no prior observed browsing history. This evaluation examines discontinuities that occur as users browse on the same device. We do not simulate the related, but different, challenge of linking users' identities across devices. Unless specified, the adversary's prior knowledge of a users browsing (i.e. number of training observations) in time period A varies ($\mu = 158.4, \sigma = 163.2$) based on the random sampling of users. Similarly, unless specified, the adversary attempts to predict the identity of the user in time period B based on a single 24-hour period of observation.

We consider this scenario to be technique-agnostic when it comes to how the browsing data was originally collected and identifiers were assigned. Our threat model assumes that an adversary has relatively wide visibility into a user's cross-site browsing activity. Such information could be collected through the use of cookies, IP-based methods via web beacons or passive network observation, or browser fingerprinting. Trackers with this ability include advertising networks, internet service providers (ISPs), DNS resolvers, or

Table 1: Comparison of previous studies on behavioral fingerprinting with our own.

Previous Studies	Year	Users	Duration	Observation Period	Data Model	Classifier	Accuracy
Yang et al. [118]	2010	100	1 year	Variable*	User Profiles	Decision Tree	62.9–87.3%
Banse et al. [9]	2012	3,000	5 months	24 hours	TF-IDF	Naive Bayes	88.2%
Herrmann et al. [46]	2013	3,600	8 weeks	24 hours	TF-IDF	Naive Bayes	85.4%
Kirchler et al. [54]	2016	3,826	8 weeks	24 hours	TF Normalized	K-Means	73.0%
Gu et al. [43]	2017	509	5 weeks	24 hours	TF-IDF	Naive Bayes	78.9%
Vassio et al. [109]	2017	2,500	4 weeks	24 hours	TF-IDF	Naive Bayes	86.6%
Bird et al. [14]	2020	19,263	2 weeks	1 week	Frequency	Jaccard Distance	50.0%
This Study	2021	100–150,000	2 years	24 hours	TF-IDF	Neural Network	36.1-85.8%

* Browsing sessions are based on user inactivity not a predefined length of time. Results are based profiles built from 1 to 100 browsing sessions.

passive observers on a network. The capabilities of these techniques and adversaries will vary. Some techniques will encounter identity discontinuities more frequently than others. Some adversaries may track a much larger population than others. Instead of attempting to model these specific scenarios, we evaluate a range of capabilities and discuss their implications as applicable.

It is important, in the context of this paper, to distinguish between the assignment of identifiers and the linking of identifiers across discontinuities. The **assignment of identifiers** is determined by the original method of tracking and constitutes what the adversary perceives as ground truth, whether accurate or not. An adversary using browser fingerprinting will inherently be less accurate in their assignment of identifiers than other methods due to 4.2–18.7% of browser fingerprints being non-unique [5, 87]. This can result in multiple users being treated as if they were all the same user. The **linking of identifiers** is a separate problem and the one that we focus on in this paper. The method of linking is successful if it can correctly associate the browsing of a given user observed in time period B with the same user, or set of users all assigned the same identifier, in time period A. Refinement of the assignment of identifiers is outside the scope of this paper, but it is another potential application of behavior-based methods.

In our analysis we vary several factors to assess their impact on the effectiveness of behavioral fingerprinting in linking user identities. First, we vary the number of users an adversary has to distinguish between to evaluate the uniqueness of fingerprints. Second, we increase the gap in time between period A and B from 0 to 540 days to assess the consistency of fingerprints as user's behavior and interests change over time. Third, we vary the number of observations an adversary has observed for each user in both time period A and B to simulate various frequencies at which discontinuities occur. Fourth, we vary the amount of activity observed in time period B to determine how quickly users lose anonymity after a discontinuity occurs and identify factors that lead users to become more or less anonymous. Finally, we examine the theoretical effectiveness of combining behavioral and browser fingerprinting to link users' identities.

3.2 Datasets

In this study we use three datasets: one new dataset for our primary evaluation and two from previous work to validate our results. The two datasets from previous work include DNS requests captured over an 8-week period in 2010 from a university network in Germany by Herrmann et al. [46] and TCP traffic collected over an 4-week period from a university network in Italy by Vassio et al. [109]. The DNS dataset collected by Herrmann et al. was kindly shared by the authors with our research team, while the TCP dataset from Vassio et al. is publicly available.[1] Our only modification to these datasets was to filter out any DNS requests in the Herrmann et al. study that were not regular name resolutions for IPv4 (DNS type A) and IPv6 (DNS type AAAA) addresses, leaving 89% of the original dataset. Filtering in this manner allows us to focus on web browsing and compare more directly to the other datasets. We will refer to these datasets as the HM and VS datasets going forward.

Our primary dataset was collected through a browser extension that provided security services to customers of a large security company in Japan.[2] The toolbar was distributed by web service partners on behalf of the security company as part of their own service offerings. The research team was not involved in the data collection process. In order to subscribe to the toolbar service, the company required its customers to agree to the terms of service and provide consent to data collection. Upon downloading the toolbar software, customers were informed that in using the security toolbar service their browsing data, including the URLs they visited, would be collected for research purposes only. If consent was not given, the toolbar software was not installed. The toolbar was only available for Microsoft Internet Explorer (IE) on desktop and laptop computers which, up until recently, had a very large user base in Japan particularly in many large businesses and government agencies where IE was the required browser [78]. We will refer to this dataset as the ST dataset.

Unlike the previous two studies which collected all DNS or TCP requests made by a user's device, data collected through the security toolbar only includes web requests generated by the web browser. However, the scale of the ST dataset far exceeds these two. Our analysis uses 25 months of data collected between January 2019 and February 2021. This subset includes data from 1,126,775 users in total, with approximately 150,000 daily active users making on average 40M requests per day. For each web request, the ST dataset

[1] Data from the Vassio et al. study is available at https://smartdata.polito.it/domains-web-users/

[2] The security company, and its browser toolbar, cannot be referred to by name due to a non-disclosure agreement with the organization.

contains a randomly-assigned unique identifier corresponding to the user, the URL visited, and a timestamp. No demographic information was collected. The research team obtained access to the ST dataset through a data-sharing agreement with the security company. Storage and processing of the shared data by the research team was conducted exclusively on secure servers. The sharing of the data received a Category 4 exemption for secondary research by the Institutional Review Board (IRB) at our academic institution(s).

3.3 Data Processing

In the following sections, we outline the steps taken to reduce the privacy risks posed by our access to the dataset, construct the feature set, and split the training and test sets.

3.3.1 Privacy. Given the privacy-sensitive nature of the data contained in the ST dataset, our first step in data processing was to minimize and mask potentially sensitive personal information. While the randomly-assigned identifiers were not linked to any direct demographic or personal information, the URLs in the dataset presented a potential privacy risk as the content of URLs, particularly URL query strings, can often contain information linked to a user's identity (e.g., usernames, emails, search terms, etc.) [115]. To address this risk, we minimized the potential exposure by stripping all of the URLs down to only the domain and subdomain information. Since domain information can also contain personally identifiable information (e.g. a personal website), we masked all of the domains and subdomains using random numeric identifiers. We then stored the mapping between the domains and identifiers on a separate secure server and restricted access to the original unmasked dataset. All subsequent data processing and analysis was performed using only the dataset of randomly assigned user identifiers, masked domain identifiers, and timestamps.

3.3.2 Feature Set. Similar to several previous studies, we model a user's browsing over a period of time as a frequency vector of the websites a user visits [9, 43, 46, 109]. We define websites at the domain level, information that is available directly to tracking using cookies or browser fingerprints or inferrable from the IP address in packet headers. The majority of previous studies generate the frequency vector using a period of 24 hours, a standard we follow and will refer to as a *browsing session*. Unlike previous work that captured DNS or TCP traffic, the ST toolbar captured web request events generated by Internet Explorer. There are a variety of events that trigger this action including the user clicking a link, using the navigation bar, refreshing a page, or other common web navigation event. As such, the unit of observation in the dataset roughly corresponds to the user navigating to a new web page. We will refer to this unit as a *request*.

We restrict the frequency vector to the 10,000 most visited websites, a point that approaches optimal accuracy while limiting the size of the model [46]. Our initial tests confirm these results, we find that increasing the size of the frequency vector to 100,000 sites improves classifier accuracy only modestly. As in previous work, we transform the frequency vector using a term frequency, inverse document frequency (TF-IDF) transformation [9, 43, 46, 109]. We recognize that the websites a user visits represents only one aspect of a user's behavior that embedded trackers could leverage for identification. Nuances in typing, scrolling, mouse movement, and clicks would provide additional identifying information and therefore our methods, in lacking this data, may underestimate what embedded trackers can achieve in practice. At the same time, hardened browsers can prevent the collection of these metrics.

3.3.3 Training and Test Sets. With the exception of some of the initial baseline tests that replicate previous methods, we generate the train and test sets in the following manner. First, we exclude sessions with fewer than five total web requests. This threshold varies in previous studies from five requests [46] to 50 requests [14]. Sessions with fewer web requests are inherently more difficult to classify as there is less information to base a decision on. We opted to use the lower end of this spectrum to 1) enable us to evaluate a greater diversity of sessions, 2) simulate realistic conditions, and 3) not over-inflate our results. Second, we only exclude a given user if they do not have at least one browsing session in the training set and one session in the test set. Several previous studies [9, 46] require a larger number of train or test sessions for some experiments, but for reasons similar to above we opt for these minimal requirements. In some tests we vary the number of users up to a maximum of 150,000—the number of average daily users contained in the dataset. For other tests, we hold the number of users constant at 25,000. This threshold was selected to maximize the number of users tested while also accounting for the limitations of our server in processing many iterative tests.

Third, since we are attempting to simulate identity discontinuities, we split the train and test sets at a specific date in time. Sessions observed on or before that date are assigned to the training set, while those observed after are assigned to the test set. This is different from previous work where a random train/test split is primarily used [9, 14, 43, 46, 109, 118]. The dates were selected to create an approximate 80/20 split between train and test sets. We split the ST dataset on October 1st 2020, the VS dataset on February 7th 2017, and the HM dataset on June 2nd 2010. We tested the ST dataset using four additional splits at four-month intervals ranging from June 1st 2019 to June 1st 2020 to ensure our selected split did not bias the results. While classifier performance declines as the split date moves further back in time, this is a natural result of having fewer training observations to learn from and more test sessions to predict. The consistency of the results across splits and lack of anomalous performance led us to conclude that our choice of date to split the data does not bias our results.

3.4 Classifier Design

Unlike previous work that has employed decision trees [118], K-Means clustering [54], Jaccard Distance [14], and Multinomial Naive Bayes [9, 43, 46, 109], we opt to use a basic feed-forward neural network. The model is composed of a single hidden layer with 10,000 nodes and a linear output layer for classification. Output of the hidden layer is batch normalized and uses a rectified linear activation function. We use a 20% dropout rate on the hidden layer output to avoid overfitting. Training was conducted using a cross-entropy loss function and an Adam optimizer.

We tuned the model and its parameters using a small validation set of 3,000 randomly selected users, equal to that tested in the Herrmann et al. study [46]. These initial tests included varying

the size of the hidden layer, the number of hidden layers, and the initial learning rate. We found that larger hidden layers provided better performance but those gains were modest after 10,000 units, a number that corresponds to the size of the feature set. Adding a second hidden layer also improved classifier performance by upwards of 5–6%, although adding subsequent layers did not effect accuracy. However, the additional layer also greatly compounded training time. Given our plans to scale the number of users (and therefore the size of the output layer) by several orders of magnitude, we opted for a single hidden layer.

3.5 Evaluation Metrics

Since our model attempts to predict the user that a given browsing session belongs to, rather than whether two sessions belong to the same user or not, our experiments do not suffer from the same class imbalance problem faced in the paper by Andriamilanto et al. [5]. Where the previous study had binary classes (match or non-match), the number of classes in our model is equal to the number of users (upwards of 150,000). That said, there is still some class imbalance due to variations in how frequently users browse the web and how long they were observed for in total. As such, we focus on reporting the F1 score of the model which is a standard metric combining of the classifier's precision and recall.

While previous work focuses on metrics related to re-identifying the exact user, we believe that it is also important to understand the effect of behavioral fingerprinting on a user's level of anonymity. As such, we also present the anonymity set for each session: the number of other users with whom the actual user is indistinguishable from. The percent of sessions with an anonymity set of at least size k is equivalent to the accuracy of the model in predicting the actual user among the top k users that a session is most likely to belong to. Therefore previous work, in solely reporting re-identification rates, provide an accuracy score only where $k = 1$. In contrast, we analyze the full distribution of anonymity sets from 1 (identifiable) to n (completely anonymous) where n is the size of the population.

4 Findings

Using our model we conduct several baseline tests to evaluate performance on the ST dataset in comparison to previous work. Next, we assess the two preconditions for behavioral fingerprints to be effective in linking identities: their uniqueness as the pool of users grows and their consistency over long periods of time. We then evaluate how the amount of behavior an adversary observes, both pre- and post-discontinuity affect their ability to link users' identities. In addition, we measure how quickly users lose anonymity after a discontinuity occurs and identify contributing factors. Finally, we assess the potential effectiveness of combining behavioral and browser fingerprinting to link users' identities.

4.1 Baseline Performance Across Datasets

The first step in our analysis is to compare the performance of our model across the ST, HM, and VS datasets. To do this we replicate the controlled test from Herrmann et al., using 18 training sessions and two test sessions for each user, assigned randomly. Since only 1,244 users met this criteria in the VS dataset, we limit our samples

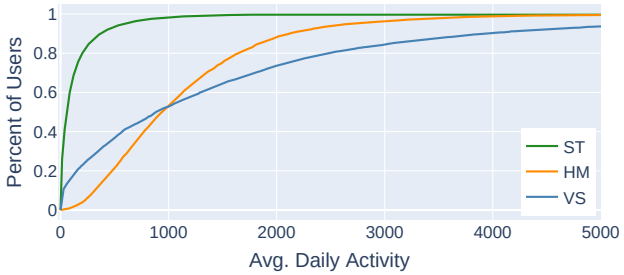


Figure 1: Cumulative distributions of users' average daily activity observed across the three datasets. Note: we restrict the y-axis to 5,000 requests, but long tails exist in all datasets.

to this number across the three datasets. We find that the performance of the neural network on the HM and VS datasets align with that reported in the previous studies. Our model achieves an 88.0% accuracy and an F1 score of 0.893 on the HM dataset, slightly higher than the 85.4% accuracy achieved by Herrmann et al.'s MNB model [46]. Similarly, the neural network has an 85.9% accuracy and F1 score of 0.875 using the VS dataset, marginally lower than the 86.6% accuracy found by Vassio et al.'s MNB model [109].

Comparatively, the neural network model performs much worse on the ST dataset achieving an accuracy of 64.7% and an F1 score of 0.654. In examining possible causes, we note that the average amount of activity observed per session is much lower in the ST dataset ($\mu = 269.3$) than in the HM ($\mu = 1,473.9$) or VS ($\mu = 3,470.4$) datasets. The differences in the distribution of average daily web activity across the three datasets can be observed in Figure 1. It is not surprising that the ST dataset captures less activity, as the toolbar only collected browser events that generated web requests (see Section 3.2 for more details). In contrast, the HM dataset contains all DNS traffic generated from the user's IP address and the VS dataset all the TCP traffic. To test the effect of this difference, we resample the VS and HM session to match the distribution found in the ST dataset. We find that this partially explains the differences in performance as the F1 score decreases from 0.893 to 0.769 on the HM dataset and from 0.875 to 0.790 on the VS dataset. Yet, this still outperforms the ST results by a fair margin.

We also hypothesize that time may also play an important role in performance. Since the ST dataset covers a period of 2 years, compared to 8 weeks for the HM dataset and 4 weeks for the VS dataset, the model using random train/test assignment on the ST dataset is likely having to make predictions across a much larger period of time. As such we switch from the random assignment used in previous work to temporal assignment as described in Section 3.4. This separates the training and test sets according to a specified date. As a secondary effect, train and test sessions tend to be closer together in time, as we require at least one train and one test session for each user. After splitting in this manner, and resampling the request distributions to match yet again, we find we find that performance between the three datasets are on par with one another. The model achieves an F1 score of 0.636 on the ST dataset, 0.680 on the VS dataset, and 0.640 on the HM dataset. Therefore, we feel confident that the browsing data collected in the ST dataset, and the results derived from it, are not anomalous. In subsequent experiments we will focus primarily on trends within

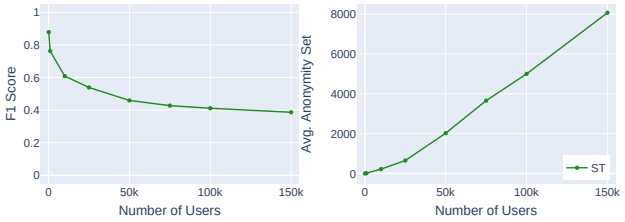


Figure 2: Changes in F1 score and average anonymity set as the number of users increase.

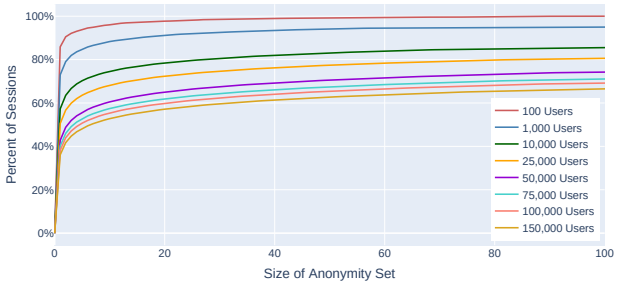


Figure 3: Cumulative distribution in the size of the anonymity for different pools of users. The x-axis is restricted to show anonymity sets smaller than 100 users.

the ST dataset, using the HM and VS datasets for validation. Absolute differences across datasets will still occur but, all else being equal, they will continue to be driven by the factors identified here.

4.2 Preconditions for Effective Linking

Since previous work has evaluated behavioral fingerprinting with small pools of users over short periods of time, one of our primary contributions is evaluating the uniqueness and consistency of behavioral fingerprints at scale. These represent the two preconditions required for effective linking across identity discontinuities.

4.2.1 Uniqueness. To assess how unique behavioral fingerprints are to an individual, the first precondition, we evaluate the scenario described in Section 3.1 with varying numbers of users. We find that while classification performance decreases with scale the marginal loss in performance due to each additional user shrinks rapidly. This is evidenced by the curve in Figure 2, which shows that the F1 score of the model drops abruptly as the size of the initial user pool grows, but then starts to level off after passing 10,000 users. As we scale beyond 50,000 users, a much larger scale than the 19,000 users tested in the largest previous study, performances stabilizes around an F1 score of 0.4. The shape of this curve is not necessarily surprising. Since the number of classes the model is trying to predict is directly proportional to the number of users, the naive baseline performance of the classification task becomes $\frac{1}{n}$ where n is the number of users. As n increases, the baseline curve would follow a similar trend, albeit one that would approach zero very rapidly.

More interestingly, average anonymity remains low even as the size of the user pool becomes very large. Also shown in Figure 2, a user's average anonymity during a given browsing session scales relatively linearly at about 5% of the total number of users. Since the addition of more users increases the number of confounding observations exponentially, this means that even when hiding among a large population of users, behavioral fingerprinting can effectively rule out the overwhelming majority of users that a given browsing session belongs to. For example, alongside 150,000 other users, one's browsing can be distinguished from 141,930 on average, reducing anonymity by 94.6%. This indicates that behavioral fingerprints are quite unique, potentially on par with browser fingerprints that are unique in 81.3–95.8% of cases [5, 87].

The risk to users is further evidenced by the cumulative distribution of anonymity sets shown in Figure 3. We restrict the x-axis in this plot to highlight anonymity sets between 0 and 100 users, demonstrating the large proportion of sessions that have little to no anonymity. Even at 150,000 users, over 55% of browsing sessions have an anonymity set less than 10 (0.007% anonymity) and over 69% less than 100 (0.067% anonymity). While there are long tails in these distributions, indicating that some sessions retain high anonymity, the majority of sessions are highly identifiable. Reducing anonymity to these levels may be sufficient for certain tracking applications, such as targeted advertising, and combined with other information may effectively link the user's identity.

In interpreting these results, remember that an adversary only needs to link the identities of users where an identity discontinuity has occurred during the same time period. This likely represents a much smaller subset than the total population that is tracked. Therefore, these methods will scale to much larger populations in total, dependent on how frequently identity discontinuities occur and how correlated they are with each other. As such, these methods will not scale as well for browser fingerprinting, where discontinuities can occur simultaneously over larger groups due to correlated software or operating system updates [63], as they would for cookie-based tracking.

4.2.2 Consistency. To test the second precondition, the consistency of behavioral fingerprints as user's habits and interests change over time, we examine how the recency of a user's fingerprint affects performance. To do this, we sample 25,000 users from the ST dataset whose browsing was consistently observed (i.e. had at least 15 sessions in each 3-month period) over the entire 2 years of data collection. In a series of experiments, we restrict the training set to sessions observed within a 90-day period, moving the window successively back in time in 90-day intervals, keeping the test set constant. The number of training sessions decreases as the sliding window moves back in time, with the largest difference over the course of 540 days being 6.3%— approximately 2.5 sessions per user. This will make the consistency of fingerprints appear slightly worse than they actually are. However, based on the results we will discuss in Section 4.3.1, the effect should be relatively small.

Despite that effect, we find that behavioral fingerprints change slowly over time. Figure 4 displays the change in F1 score and average anonymity as the gap between training observations and test sessions increases from 0 to 540 days. Over this period, we find that the F1 score decreases from 0.571 to 0.339 and the size of the average anonymity set increases from 422.3 (1.7%) to 1,783.5 (7.1%). In both cases, the trend is linear with a very gradual slope. This indicates that changes in a user's interests do not completely alter their day-to-day online behavior and that larger shifts in behavioral patterns occur slowly over the course of months to years. Most

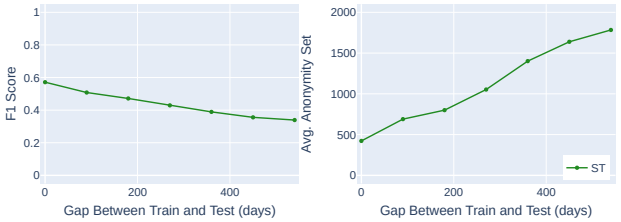


Figure 4: Change in F1 score and average anonymity set as the gap in time between training and test sets increases from 0 to 540 days.

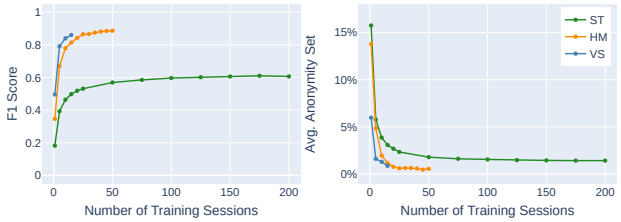


Figure 5: Changes in F1 score and average anonymity as the number of training sessions per user increase.

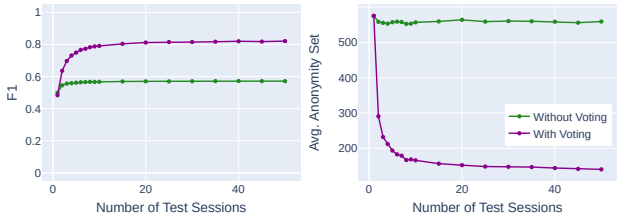


Figure 6: Changes in F1 score and average anonymity set using weighted voting varying the number of test sessions per user contributing to the vote.

importantly, the rate of change is much slower than that of browser fingerprints, which can change substantially over the course of several days or even hours [5, 63, 111]; IP addresses, which are retained for an average of 19 days in the United States [73]; and even browser cookies, of which 60% are reset within a month [29]. Together with the findings in Section 4.2.1, these results confirm the findings of previous work [28, 31, 83] and provide further evidence that individuals' behavioral fingerprints are unique and consistent, satisfying the preconditions for effective identity linking.

4.3 The Effect of Observation Length

Intuitively, adversaries who observe a user's browsing behavior for longer periods of time are more likely to pick up on their habitual patterns. In the following sections, we explore how much information, both before and after an identity discontinuity occurs, an adversary needs to identify those patterns and link users' identities. In both cases, the amount of information available is largely dependent on the frequency at which identity discontinuities occur.

4.3.1 Pre-Discontinuity. To assess the effect of an adversary's prior knowledge of user's web browsing, we run a series of experiments varying the number of training sessions available per user. For the ST dataset, these tests were run with 25,000 users who could each provide at least 200 training sessions. We validate the results with 2,000 users across 15 and 50 sessions for the VS and HM datasets respectively. Figure 5 displays changes in F1 score and average anonymity as the number of training sessions increase across the three datasets. While performance is initially quite low, F1 scores reach near-maximal levels with 15–25 observed sessions. Only marginal improvements are gained after 50 training sessions. Although some curves are truncated due to the smaller size of the dataset, we observed the same trend across all three. The size of the average anonymity set drops rapidly until 15–25 sessions are observed after which changes are relatively small.

The rates at which IP addresses, with an average retention of 19 days in the U.S. [73], and browser cookies, where 30% are reset monthly and 40% are retained longer [29], tend to change fall within or beyond this range. In contrast, browser fingerprints are likely to change before an adversary observes 15–25 sessions. However, our results demonstrate that even with only one prior observation (i.e. one training example) per user, average anonymity is reduced by 84–95% depending on the dataset. That means that even with minimal prior knowledge, an adversary can eliminate most of the anonymity gains from an identity discontinuity occurring.

4.3.2 Post-Discontinuity. To evaluate how the amount of information an adversary observes post-discontinuity affects performance, we introduce a method of *weighted voting*. Instead of predicting the user individually for each test session as we have done in the previous experiments, this method combines the predicted probabilities across multiple test sessions belonging to the same user. Assuming that a user's habitual patterns are more likely to be revealed during browsing sessions with greater activity, we apply a weight to the probabilities based on the number of observed requests in the session. The user with the largest sum of weighted probabilities is selected as the prediction for all sessions. We evaluate this weighted voting method using a fixed set of 25,000 users that can each contribute at least 50 test sessions. We then vary the number of test sessions per user while maintaining a fixed training set.

Figure 6 shows the effect of the number of test sessions on F1 score and the average anonymity set size for the model with weighted voting (purple) and without (green). Weighted voting increases model performance considerably, reaching a maximum F1 score of 0.821 at 50 test sessions. That is a 43.6% increase over the same model without weighted voting. At this point, the classifier can correctly identify the user in 71.5% of cases and for 87.9% of sessions can reduce the size of the anonymity set to at most 25 users (a reduction of 99.9%). We find that the gains in performance level off after an adversary has observed 10 sessions post-discontinuity. Like before, this indicates that behavioral fingerprinting is more effective in linking identifiers that are stable for longer periods of time, such as cookies and IP addresses. However, our results show there are diminishing returns for each additional day observed. The largest gains are found having just observed one additional session per user, resulting in a 27.7% increase to the model's F1 score and a 49.4% reduction in average anonymity. This demonstrates that behavioral fingerprinting remains relevant even when discontinuities occur more frequently, as is the case with browser fingerprinting.

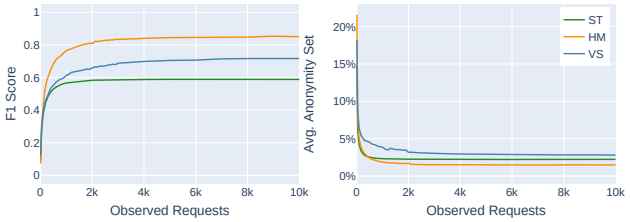


Figure 7: F1 score and average anonymity as amount of observed activity in a session grows.

4.4 Anonymity Loss after a Discontinuity

Next we turn to an evaluation of how users, who after a discontinuity are anonymous, lose their anonymity as the spend time online. We first examine how anonymity declines with the amount of activity a tracker observes. Then we identity factors affecting whether users retain their anonymity.

4.4.1 Anonymity Loss. We know from the baseline evaluation in Section 4.1 that the amount of user activity observed has a substantial impact on performance between the three datasets. In this experiment we limit the amount of observed activity per test session starting with 5 requests, simulating when a user first comes online, and extending it to 10,000 requests, which covers most browsing over the course of a week. Sessions in the training set contain all observed requests and therefore remain constant throughout all runs. Like before, we use a sample of 25,000 users in the ST dataset and 2,000 users in the VS and HM datasets.

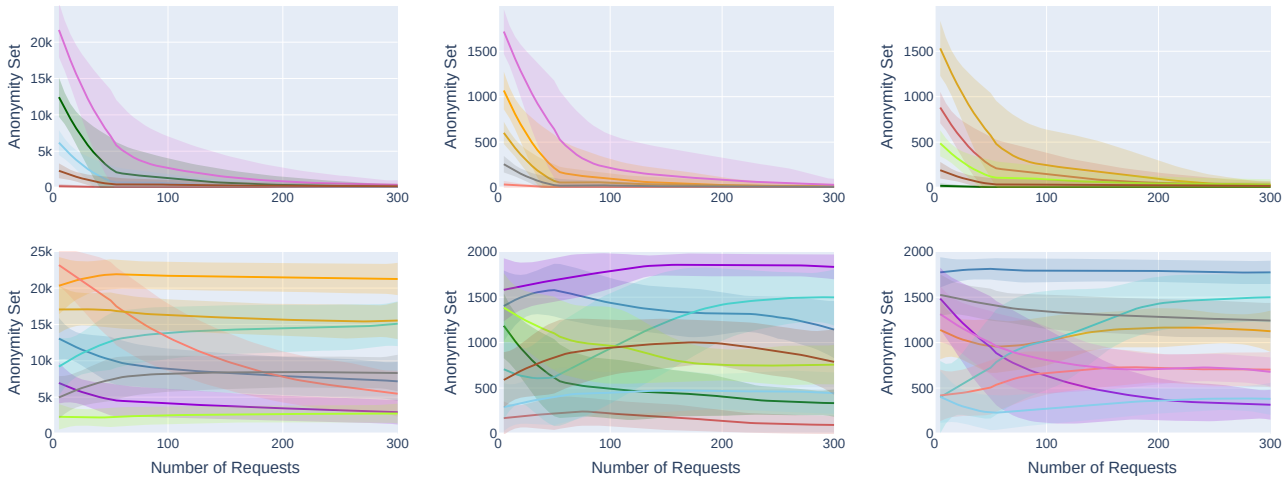
Figure 7 shows the change in F1 score and average anonymity as the number of observed requests increases. While it takes 1,000–2,000 requests to maximize classifier performance, anonymity plummets almost as soon as a user begins browsing. A user loses between 78–85% of their anonymity on average within the first five requests the tracker observes, more than 90% within 30 requests, and over 95% within 100–250 requests. Based on our three datasets, 5 requests are typically made within the first 15–60 seconds of browsing, 30 requests within the first 3.5–10 minutes, and 100–250 requests between 15–50 and 65–160 minutes depending on the dataset. This has two major implications. First, behavioral fingerprinting may enable a tracker to perform rapid linking among groups of users, even if they cannot at the individual level. Given that those groups have similar browsing patterns and visit the same sets of websites, this may be sufficient for some tracking purposes, such as targeted advertising. Second, this greatly reduces the effectiveness of privacy protections based on inducing identity discontinuities, unless done at a very frequent rate. Such changes may negate the utility of intelligence tracking protection [112] and similar browser-based protections [27] that try to strike a balance between retaining cookies for usability and clearing them for privacy.

4.4.2 Factors Affecting Anonymity Loss. Our results up until this point indicate that for many sessions an adversary is able to accurately narrow in on who the user is, but for a minority of sessions they would have no sense of the user's identity—so what is different about these types of sessions? To answer this question we disaggregate the previous results to detect common patterns in the anonymity curves of different sessions. We hesitate to arbitrarily designate a cutoff for when a session's anonymity is sufficiently reduced to designate the user as identifiable—is it a single user, two, ten, a hundred? As such, we use unsupervised clustering methods, specifically time-series k-means with dynamic time warping, to extract trends. We apply a Savitzky–Golay filter to smooth the curves as a preprocessing step and rerun the clustering process multiple times, using 2 to 20 clusters, which allows the patterns to emerge naturally until no new trends can be identified.

Using these methods, we find that beyond 13 unique clusters no new trends emerge. Figure 8 displays the results across the three datasets. Each curve represents the average anonymity of a given cluster and the shaded area one standard deviation. These clusters are separated into two distinct groups: identifiable curves where anonymity approaches zero on top, and anonymous curves where anonymity remains high on the bottom. As illustrated in the subplots, the patterns that emerge during clustering are consistent across the ST, HM, and VS datasets.

We find that for the average user, the majority of their browsing sessions fall into one of the identifiable clusters. On average, 81.0% of a given user's sessions in the ST dataset become identifiable, 88.6% in the VS dataset, and 90.2% in the HM dataset. However, most users also have at least one session that remains completely anonymous. To understand what differentiates these types of sessions, we employ a logistic regression whose dependent variable represents whether the session was in the identifiable clusters or one of the anonymous sets. We run the logistic regression model on all three datasets. The independent variables we test are summarized in Table 2 alongside the regression results. They include the number of requests observed, the unique domains visited by the user, the average traffic rank of sites visited, the number of days between the last training session ended and the observed session, and the number of training sessions available for that user. The last two independent variables, the maximum site prevalence ratio and bigram prevalence ratio, represent how closely related the most unique site, or combination of two sites, is to the user as compared to the rest of the population. The prevalence ratio, a common metric in epidemiology, in our work refers the proportion of browsing a specific user spends visiting a given website compared to the proportion of browsing that all users spend visiting that site. As such, higher prevalence ratios is a measure of how niche a user's browsing is. Several variables exhibiting log-normal distributions were log-transformed for the regression.

Table 2 shows fairly consistent results across the three datasets. We find that visiting a greater number of pages (requests) and frequently returning to niche websites (site prevalence ratio), a signal of unique interests, are positively correlated with anonymity loss, although the former is not statistically significant in the HM dataset. This is indicated by an odds ratio, derived from the regression coefficients, that is substantially greater than 1. The interpretation of the odds ratio is as follows; a 1% increase in the number of requests observed is associated with being 1.2–1.7 times more likely that the session will become identifiable depending on the dataset. In contrast, odds ratios substantially less than 1 indicate behaviors that are positively correlated with anonymity. Browsing more popular sites (those with a high traffic rank) and a greater diversity of sites, which can potentially mask unique habits or interests, tend to keep a user more anonymous.



(a) ST Dataset (b) VS Dataset (c) HM Dataset

Figure 8: Unsupervised clustering of anonymity curves as the number of observed requests within a session increases.

Table 2: Logistic regression regressing session identifiability on metrics reflecting variation in user browsing behavior.

	ST				VS				HM			
Variable	**Coeff.**	**Error**	***p*-value**	**Odds Ratio**	**Coeff.**	**Error**	***p*-value**	**Odds Ratio**	**Coeff.**	**Error**	***p*-Value**	**Odds Ratio**
Intercept	-0.674*	0.052	< 0.001	0.510	0.894	0.625	0.152	2.445	7.630*	1.428)	< 0.001	2059.874
Number of Requests (Log)	0.223*	0.008	< 0.001	1.250	0.558*	0.110	< 0.001	1.747	0.185	0.206	0.370	1.203
Unique Domains Visited (Log)	-0.3619*	0.012	< 0.001	0.0696	-0.910*	0.113	< 0.001	0.402	-0.447*	0.168	0.008	0.640
Avg. Traffic Rank (Log)	-0.233*	0.006	< 0.001	0.792	-0.216*	0.077	0.005	0.805	-0.664*	0.127	< 0.001	0.515
Days After Training Cutoff	-0.005*	< 0.001	< 0.001	0.995	-0.022	0.012	0.071	0.979	-0.022*	0.006	< 0.001	0.978
Number of Training Sessions	0.034*	< 0.001	< 0.001	1.034	0.0176*	0.003	< 0.001	1.018	0.0016	0.004	0.647	1.002
Max Site Prevalence Ratio (Log)	0.400*	0.009	< 0.001	1.492	0.531*	0.060	< 0.001	1.701	0.460*	0.071	< 0.001	1.584
Max Bigram Prevalence Ratio (Log)	0.051*	0.007	< 0.001	1.052	-0.054	0.08	0.504	0.948	-0.166	0.122	0.176	0.847

* Statistically significant at the 1% confidence level.

The remaining explanatory variables have a very small effect, demonstrated by an odds ratio close to 1. Surprisingly, this includes the maximum bigram prevalence ratio. We hypothesized that pairs of sites that are relatively unique to the user would be an important driver of anonymity loss, yet that is not the case. One explanation is that specific websites are closely associated with individual users. However, this seems unlikely given we only use the top 10,000 sites in the feature vector yet the classifier is able to distinguish between an order of magnitude more users fairly well. An alternative hypothesis is that our bigram metric captures pairs of visits that are unique to the user and to the specific session, providing no information to link different sessions belonging to the same user.

4.5 Behavioral and Browser Fingerprinting

Up to this point, we have evaluated the ability of behavioral fingerprinting to link users' identities across discontinuities in isolation. However, as we discussed in Section 2.4, similarities in browser fingerprints can provide some, albeit limited, capability to link identities, even as they change rapidly over time [5, 63, 87, 111]. Intuitively, an adversary with access to a greater number of data points about an individual, should be able to pick them out of a crowd more easily. In this section, we assess how an adversary might leverage the combination of behavioral and browser fingerprinting. Since the ST dataset does not contain users' browser fingerprints, we base our analysis on the combination of our results and that of Andriamilanto et al.'s study on browser fingerprinting [5]. In doing so, we rely on the assumption that a user's browsing behavior and device configuration are largely independent. Although likely to be true, we discuss the limitations further in Section 5.3.

To employ this combined approach, an adversary could 1) employ cookies or IP addresses as the primary means of tracking and collect browser fingerprints, such as for the purpose of fingerprint-based cookie respawning [39], or 2) use browser fingerprinting as the sole means of tracking. As discussed in Section 3.1, the assignment of identifiers in the latter case suffers from an underlying error rate due to non-unique fingerprints. However, like before, we focus on the effectiveness of linking identifiers, not on the accuracy of their original assignment. When an identity discontinuity occurs, an adversary using both fingerprinting techniques will have immediate access to the anonymous users' new browser fingerprint. As the anonymous user browses over time, the adversary will begin to piece together their behavioral fingerprint. In the following sections, we first simulate an adversary iteratively refining the linking process over time after a discontinuity occurs. We then assess how the number of users affect these methods at scale.

4.5.1 Iterative Approach. Immediately after an identity discontinuity, the only data available to an adversary are browser fingerprints.

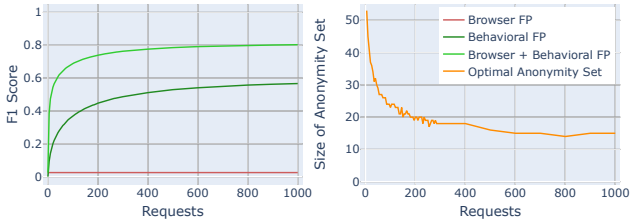


Figure 9: Comparison of fingerprinting techniques as an adversary iteratively builds behavioral fingerprints over time.

At this point, they could employ the linking method proposed by Andriamilanto et al. that achieves a 0.61% false positive and negative rate [5]. Given the much larger number of comparisons between fingerprints from different browsers as discussed in Section 2.4, a false positive rate of 0.61% generates a large number of false positives. With 25,000 users these methods produce 78 false positives for every one true positive. Although recall (0.994) is high due to the small number of false negatives, precision (0.013) is extremely low. Overall, we find that these methods yield an F1 score of 0.025.

As time goes on, the adversary will start to collect, and then continue to refine, users' behavioral fingerprints. Using behavioral fingerprints, an adversary can reduce the anonymity set associated with each observed browsing session and then compare browser fingerprints among the smaller pool of potential users, greatly reducing the number of false positives and improving precision. Figure 9 shows the performance of the browser fingerprinting approach (in red), the behavioral fingerprinting approach (in dark green), and the combined approach (in light green) as an adversary observes more user activity (requests). First, this illustrates that the combined fingerprinting approach performs better in the long run, approaching F1 scores of 0.81. Second, this method can link users' identities more rapidly and with greater accurately, reaching an F1 score of 0.58 within the first 30 requests observed or, alternatively, 3.5–10 minutes of browsing. We find that the size of the anonymity set can be dynamically optimized during this iterative process to achieve maximum performance. Also shown in Figure 9, an adversary with limited information should use larger anonymity sets, with upwards of 53 users being optimal. As more browsing data is gathered, the optimal size decreases until reaching 13 users.

4.5.2 At Scale. Using the same iterative approach, we assess how these methods perform across a varying number of users. For comparison, we only report the final results where an adversary has complete fingerprinting information. The results of our analysis are summarized in Figure 10 which shows the F1 score across different numbers of users for Andriamilanto et al.'s methods [5] (red), our behavioral fingerprinting models with (light purple) and without weighted voting (light green), and the combined approach with (dark purple) and without weighted voting (dark green).

As illustrated, the performance of the browser fingerprinting method alone scales very poorly with F1 scores dropping below 0.4 for 1,000 users and below 0.1 for 10,000 users. On their own, the behavioral fingerprinting methods scale much better. Without weighted voting, F1 scores stay above 0.38 as users approach 150,000. With weighted voting, F1 scores remain above 0.67 as users approach 100,000. Note that the approach using weighted

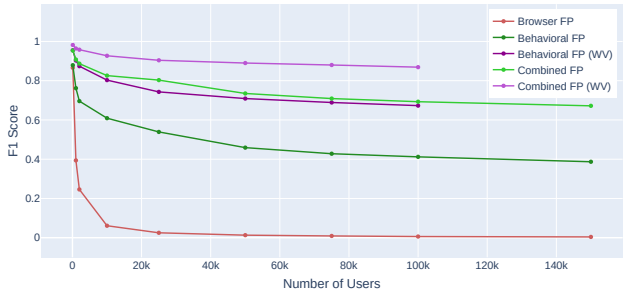


Figure 10: Comparison of browser fingerprinting, behavioral fingerprinting, and the combination of techniques at scale.

voting was only tested with up to 100,000 users due to server resource constraints. Overall, the combined approach outperforms each method individually achieving F1 scores above 0.67 without weighted voting and above 0.86 with weighted voting.

These results indicate that an adversary, in combining fingerprinting techniques, can link users' identities rapidly and effectively at scale. However, these estimates are theoretical in nature, combining our experimental results with the reported findings of previous work [5]. Future work should examine this combination empirically, particularly to assess how an adversary may use both fingerprints in concert (i.e. use a joint feature set in a single classification step) and how correlated users' behavioral and browser fingerprints are.

5 Discussion

We now discuss the implications of work for online privacy, identify areas for future work, and highlight the limitations of our work.

5.1 Ramifications for Privacy

Our results provide further evidence that users' behavioral fingerprints are both unique and consistent, prerequisites for linking identities across discontinuities. In addition, we show that behavioral fingerprinting can be very effective in reducing user anonymity online. This remains true at scale and with limited observation. As such, behavioral fingerprinting presents a viable solution to the identity discontinuity problem, potentially enabling adversaries to conduct persistent long-term tracking.

After a discontinuity occurs, we show that behavioral fingerprinting can greatly reduce a user's anonymity as they start browsing again. Combining fingerprinting techniques can accelerate this anonymity loss even further. This greatly degrades the effectiveness of privacy-preserving techniques that induce identity discontinuities. However, our results indicate that there are two factors that privacy-preserving technologies can leverage in order to improve defenses against these techniques. The first is increasing the frequency at which actions like deleting cookies, changing IP address, or modifying the attributes of browser fingerprints occur. Shortening the time period in which any single identifier can be used to track an individual, reduces the amount of behavioral information an adversary could use for linking purposes. The second is to coordinate discontinuities across users so they are induced concurrently, thereby increasing the pool of confounding users.

While behavioral fingerprinting can be inherently conducted alongside any form of web tracking, in combination with browser

fingerprinting additional privacy risks are raised. Although behavioral fingerprinting does not solve the problem of non-unique browser fingerprints, it can aid in linking together browser fingerprints as they change over time—a critical limitation at scale [63]. This can potentially lead to more robust stateless cross-site tracking that is difficult to detect and therefore harder to protect against [68]. The lack of detection complicates the enforcement of consent mechanisms required by GDPR and CCPA—potentially enabling adversaries to circumvent these protections entirely.

Raising further privacy concerns, these techniques could also be adopted by law enforcement and government authorities. Under third-party record doctrine in the United States, established in United States v. Miller (1976) [103] and Smith v. Maryland (1979) [104], law enforcement can access browsing data collected by private companies without a warrant. Furthermore, under the Electronic Communications Privacy Act of 1986 law enforcement can obtain the IP addresses of the websites a user visits using a subpeona that does not require a judge's approval [102]. With access to this data, law enforcement and government agencies could use behavioral fingerprinting to potentially trace user activity over long periods of time, narrowing down or identifying persons of interest based on their web browsing. This can lead to positive outcomes and may be particularly pertinent to certain digital crimes. However, in regions with repressive laws these same techniques could be employed to further surveillance of citizens and enforce censorship.

5.2 Future Work

To improve user privacy, we recommend that future research explore the following five areas. First, researchers should examine the combined performance of behavioral and browser fingerprinting in practice. Second, efforts to investigate ways to better detect and prevent the collection of fingerprinting attributes should be continued. These efforts will be critical for enforcement of existing privacy regulation that relies on notice and choice. Third, further research into obscuring behavioral fingerprints, like the use of injected traffic, should be prioritized. The findings in Section 4.4.2 suggest that injecting small amounts of traffic to random niche sites, signaling interests unrelated to the user, could help maintain anonymity. Fourth, additional behavioral attributes, such as a user's activity within a web page, should be examined to further estimate potential fingerprinting capabilities. Finally, while we did not examine cross-device tracking or disaggregating the web activity of multiple anonymous users, our results provide evidence that behavioral fingerprinting could be effective for such purposes. Examples include the disaggregation of multiple users web traffic originating from a single IP address or proxy server, the identification of previously unobserved users, and the disentanglement or refinement of confounding identifiers (e.g. non-unique browser fingerprints). Future work should explore these potential threats.

5.3 Limitations

There are several important limitations to our work. First, while we find similar trends across the three datasets examined in our study, there may be other contexts where our results do not hold. The fact that these datasets were collected in countries that speak different languages (Japanese, German, and Italian), have distinct cultures, and likely visit different (but overlapping) sets of websites speaks to the robustness of our results. German, Japanese, and Italian rank as the second, fourth, and eighth most common languages on the internet by web page count [26]. As such we expect our findings to be consistent in English-speaking countries, as English language websites comprise the largest proportion of the internet and therefore have a theoretically higher upper bound on behavioral fingerprint uniqueness. However, for populations who speak languages with a smaller internet footprint, behavioral fingerprints may be much less unique. In addition, the ST dataset was limited to browsing conducted using Internet Explorer, likely in work environments, on personal computers. While the consistent results found using the VS and HM datasets—which were collected using browser-agnostic methods in university environments—provide evidence that our results from the ST dataset generalize beyond these contexts, further validation is required particularly with regard to mobile browsing.

Second, in our experiments we require users to have at least one training and one test session to be included in the sample. However, in a real world environment there may be users who have been observed in the past but do appear among the sessions being classified. Similarly, there may be new users who appear among the prediction set for whom tracker has no prior history. Both cases will likely degrade classifier performance. Third, since we do not have browser fingerprints for the users in the three datasets we examined, our performance estimates for combined browser/behavioral techniques in Section 4.5 are based on reported results from previous work. We cannot confirm whether the results from previous work would generalize to the datasets we use. Fourth, our analysis of the combined approach assumes that a user's behavioral fingerprint and browser fingerprint are independent. It is likely that these fingerprints are not strongly correlated with each other and, therefore, are more unique in combination than they are individually. However, there may be aspects of a device's configuration, such as the choice of web browser, that affect users' behavioral and browser fingerprints in similar ways. As such, our calculations may inflate the potential of combining these techniques. Fifth, while we exclude sessions with less than five requests, which we believe is a reasonable (if not overly conservative) assumption to make in practice, this does inflate our results compared to including all possible sessions. Sixth, since we did not have access to any demographic information about participants, we do not know if the model performs differently across various subgroups.

6 Conclusion

In this paper we demonstrate how an adversary can use behavioral fingerprinting to potentially overcome the identity discontinuity problem, thereby enabling persistent, privacy-invasive long-term tracking. We find that behavioral techniques are very effective in reducing the anonymity of users even when an adversary is operating with outdated and limited information. We demonstrate that users lose anonymity very rapidly after an identity discontinuity occurs, potentially limiting the effectiveness of some privacy-preserving techniques, and identify key factors that make users more or less anonymous. Finally, we show how the combination of behavioral and browser fingerprinting can potentially be combined to provide even more effective linking capabilities.

Acknowledgments

The authors would like to thank Dominik Herrmann, Christian Banse, and Hannes Federrath—authors of the paper entitled "Behavior-based Tracking: Exploiting Characteristic Patterns in DNS Traffic"—for sharing their dataset with us (the HM Dataset). This research was partially funded by the National Security Agency (NSA) Science of Security Lablet at Carnegie Mellon University (contract #H9823014C0140) and by a gift from KDDI Research, Inc.

References

[1] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. 2014. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security* (Scottsdale, Arizona, USA) *(CCS '14)*. Association for Computing Machinery, New York, NY, USA, 674–689. https://doi.org/10.1145/2660267.2660347

[2] Lalit Agarwal, Nisheeth Shrivastava, Sharad Jaiswal, and Saurabh Panjwani. 2013. Do Not Embarrass: Re-Examining User Concerns for Online Tracking and Advertising. In *Proceedings of the Ninth Symposium on Usable Privacy and Security* (Newcastle, United Kingdom) *(SOUPS '13)*. Association for Computing Machinery, New York, NY, USA, Article 8, 13 pages. https://doi.org/10.1145/2501604.2501612

[3] Nasser Mohammed Al-Fannah, Wanpeng Li, and Chris J. Mitchell. 2018. Beyond Cookie Monster Amnesia: Real World Persistent Online Tracking. In *Information Security*, Liqun Chen, Mark Manulis, and Steve Schneider (Eds.). Springer International Publishing, Cham, 481–501.

[4] Muhammad Ali, Piotr Sapiezynski, Miranda Bogen, Aleksandra Korolova, Alan Mislove, and Aaron Rieke. 2019. Discrimination through optimization: How Facebook's ad delivery can lead to skewed outcomes. *CoRR* abs/1904.02095 (2019). arXiv:1904.02095 http://arxiv.org/abs/1904.02095

[5] Nampoina Andriamilanto, Tristan Allard, Gaëtan Le Guelvouit, and Alexandre Garel. 2021. A Large-scale Empirical Analysis of Browser Fingerprints Properties for Web Authentication. *ACM Transactions on the Web* 16, 1 (sep 2021), 1–62. https://doi.org/10.1145/3478026

[6] Julia Angwin and Terry Parris. 2016. Facebook lets advertisers exclude users by race. https://www.propublica.org/article/facebook-lets-advertisers-exclude-users-by-race

[7] Julia Angwin, Noam Scheiber, and Ariana Tobin. 2017. Dozens of Companies Are Using Facebook to Exclude Older Workers From Job Ads. https://www.propublica.org/article/facebook-ads-age-discrimination-targeting

[8] Mika Ayenson, Dietrich Wambach, Ashkan Soltani, Nathaniel Good, and Chris Hoofnagle. 2011. Flash cookies and privacy II: now with HTML5 and ETag respawning. *SSRN Electronic Journal* (07 2011). https://doi.org/10.2139/ssrn.1898390

[9] Christian Banse, Dominik Herrmann, and Hannes Federrath. 2012. Tracking Users on the Internet with Behavioral Patterns: Evaluation of Its Practical Feasibility. In *27th Information Security and Privacy Conference (SEC) (Information Security and Privacy Research, Vol. AICT-376)*, Dimitris Gritzalis, Steven Furnell, and Marianthi Theoharidou (Eds.). Springer, Heraklion, Crete, Greece, 235–248. https://doi.org/10.1007/978-3-642-30436-1_20 Part 6: Privacy Attitudes and Properties.

[10] Paul Barford, Igor Canadi, Darja Krushevskaja, Qiang Ma, and S. Muthukrishnan. 2014. Adscape: Harvesting and Analyzing Online Display Ads. In *Proceedings of the 23rd International Conference on World Wide Web* (Seoul, Korea) *(WWW '14)*. Association for Computing Machinery, New York, NY, USA, 597–608. https://doi.org/10.1145/2566486.2567992

[11] Muhammad Ahmad Bashir and Christo Wilson. 2018. Diffusion of User Tracking Data in the Online Advertising Ecosystem. *Proceedings on Privacy Enhancing Technologies* 2018 (2018), 103 – 85.

[12] Peter Baumann, Stefan Katzenbeisser, Martin Stopczynski, and Erik Tews. 2016. Disguised Chromium Browser: Robust Browser, Flash and Canvas Fingerprinting Protection. In *Proceedings of the 2016 ACM on Workshop on Privacy in the Electronic Society* (Vienna, Austria) *(WPES '16)*. Association for Computing Machinery, New York, NY, USA, 37–46. https://doi.org/10.1145/2994620.2994621

[13] Reuben Binns, Jun Zhao, Max Van Kleek, and Nigel Shadbolt. 2018. Measuring Third-Party Tracker Power across Web and Mobile. *ACM Trans. Internet Technol.* 18, 4, Article 52 (Aug 2018), 22 pages. https://doi.org/10.1145/3176246

[14] Sarah Bird, Ilana Segall, and Martin Lopatka. 2020. Replication: Why We Still Can't Browse in Peace: On the Uniqueness and Reidentifiability of Web Browsing Histories. In *Sixteenth Symposium on Usable Privacy and Security (SOUPS 2020)*. USENIX Association, 489–503. https://www.usenix.org/conference/soups2020/presentation/bird

[15] Károly Boda, Ádám Máté Földes, Gábor György Gulyás, and Sándor Imre. 2011. User Tracking on the Web via Cross-Browser Fingerprinting. In *Nordic Conference on Secure IT Systems*.

[16] Elijah Robert Bouma-Sims, Megan Li, Yanzi Lin, Adia Sakura-Lemessy, Alexandra Nisenoff, Ellie Young, Eleanor Birrell, Lorrie Faith Cranor, and Hana Habib. 2023. A US-UK Usability Evaluation of Consent Management Platform Cookie Consent Interface Design on Desktop and Mobile. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI '23)*. Association for Computing Machinery, New York, NY, USA, Article 163, 36 pages. https://doi.org/10.1145/3544548.3580725

[17] Aaron Cahn, Scott Alfeld, Paul Barford, and S. Muthukrishnan. 2016. An Empirical Study of Web Cookies. In *Proceedings of the 25th International Conference on World Wide Web* (Montréal, Québec, Canada) *(WWW '16)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 891–901. https://doi.org/10.1145/2872427.2882991

[18] California Legislative Bill. 2018. California Consumer Privacy Act of 2018.

[19] Canvas Defender. 2017. Canvas Defender - Firefox Add-on that Adds Unique and Persistent Noise to a Canvas Element. https://addons.mozilla.org/en-US/firefox/addon/no-canvas-fingerprinting/

[20] Yinzhi Cao, Zhanhao Chen, Song Li, and Shujiang Wu. 2017. Deterministic Browser. In *Proceedings of the 2017 ACM SIGSAC Conference on Computer and Communications Security* (Dallas, Texas, USA) *(CCS '17)*. Association for Computing Machinery, New York, NY, USA, 163–178. https://doi.org/10.1145/3133956.3133996

[21] Yinzhi Cao, Song Li, and Erik Wijmans. 2017. (Cross-)Browser Fingerprinting via OS and Hardware Level Features. In *Network and Distributed Systems Security Symposium (NDSS)*.

[22] Juan Miguel Carrascosa, Jakub Mikians, Ruben Cuevas, Vijay Erramilli, and Nikolaos Laoutaris. 2015. I Always Feel like Somebody's Watching Me: Measuring Online Behavioural Advertising. In *Proceedings of the 11th ACM Conference on Emerging Networking Experiments and Technologies* (Heidelberg, Germany) *(CoNEXT '15)*. Association for Computing Machinery, New York, NY, USA, Article 13, 13 pages. https://doi.org/10.1145/2716281.2836098

[23] Darion Cassel, Su-Chin Lin, Alessio Buraggina, William Wang, Andrew Zhang, Lujo Bauer, Hsu-Chun Hsiao, Limin Jia, and Timothy Libert. 2022. OmniCrawl: Comprehensive Measurement of Web Tracking With Real Desktop and Mobile Browsers. *Proceedings on Privacy Enhancing Technologies* 2022 (01 2022), 227–252. https://doi.org/10.2478/popets-2022-0012

[24] Farah Chanchary and Sonia Chiasson. 2015. User Perceptions of Sharing, Advertising, and Tracking. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa, 53–67. https://www.usenix.org/conference/soups2015/proceedings/presentation/chanchary

[25] Claude Chaudet, Eric Fleury, Isabelle Guérin Lassous, Hervé Rivano, and Marie-Emilie Voge. 2005. Optimal positioning of active and passive monitoring devices. In *Proceedings of the 2005 ACM Conference on Emerging Network Experiment and Technology* (Toulouse, France) *(CoNEXT '05)*. Association for Computing Machinery, New York, NY, USA, 71–82. https://doi.org/10.1145/1095921.1095932

[26] Common Crawl. 2023. Statistics of Common Crawl Monthly Archives: Distribution of Languages. https://commoncrawl.github.io/cc-crawl-statistics/plots/languages

[27] cookiestatus.com. 2025. Cookie Status. https://www.cookiestatus.com/

[28] Kyle Crichton, Nicolas Christin, and Lorrie Faith Cranor. 2021. How Do Home Computer Users Browse the Web? *ACM Trans. Web* 16, 1, Article 3 (sep 2021), 27 pages. https://doi.org/10.1145/3473343

[29] Anirban Dasgupta, Maxim Gurevich, Liang Zhang, Belle Tseng, and Achint O. Thomas. 2012. Overcoming browser cookie churn with clustering. In *Proceedings of the Fifth ACM International Conference on Web Search and Data Mining* (Seattle, Washington, USA) *(WSDM '12)*. Association for Computing Machinery, New York, NY, USA, 83–92. https://doi.org/10.1145/2124295.2124308

[30] Amit Datta, Michael Carl Tschantz, and Anupam Datta. 2014. Automated Experiments on Ad Privacy Settings: A Tale of Opacity, Choice, and Discrimination. *CoRR* abs/1408.6491 (2014). arXiv:1408.6491 http://arxiv.org/abs/1408.6491

[31] Clemens Deußer, Steffen Passmann, and Thorsten Strufe. 2020. Browsing Unicity: On the Limits of Anonymizing Web Tracking Data. In *2020 IEEE Symposium on Security and Privacy (SP)*. 777–790. https://doi.org/10.1109/SP40000.2020.00018

[32] Peter Eckersley. 2010. How Unique is Your Web Browser?. In *Proceedings of the 10th International Conference on Privacy Enhancing Technologies* (Berlin, Germany) *(PETS'10)*. Springer-Verlag, Berlin, Heidelberg, 1–18.

[33] Serge Egelman and Eyal Peer. 2015. The Myth of the Average User: Improving Privacy and Security Systems through Individualization. In *Proceedings of the 2015 New Security Paradigms Workshop* (Twente, Netherlands) *(NSPW '15)*. Association for Computing Machinery, New York, NY, USA, 16–28. https://doi.org/10.1145/2841113.2841115

[34] Steven Englehardt and Arvind Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of ACM CCS 2016*.

[35] Steven Englehardt, Dillon Reisman, Christian Eubank, Peter Zimmerman, Jonathan Mayer, Arvind Narayanan, and Edward W. Felten. 2015. Cookies That

Give You Away: The Surveillance Implications of Web Tracking. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 289–299. https://doi.org/10.1145/2736277.2741679

[36] European Commission. 2016. Regulation (EU) 2016/679 of the European Parliament and of the Council of 27 April 2016 on the protection of natural persons with regard to the processing of personal data and on the free movement of such data, and repealing Directive 95/46/EC (General Data Protection Regulation) (Text with EEA relevance). https://eur-lex.europa.eu/eli/reg/2016/679/oj

[37] Amin FaizKhademi, Mohammad Zulkernine, and Komminist Weldemariam. 2015. FPGuard: Detection and Prevention of Browser Fingerprinting. In *IFIP Annual Conference on Data and Applications Security and Privacy*. 293–308. https://doi.org/10.1007/978-3-319-20810-7_21

[38] Ugo Fiore, Aniello Castiglione, Alfredo De Santis, and Francesco Palmieri. 2014. Countering Browser Fingerprinting Techniques: Constructing a Fake Profile with Google Chrome. In *2014 17th International Conference on Network-Based Information Systems*. 355–360. https://doi.org/10.1109/NBiS.2014.102

[39] Imane Fouad, Cristiana Santos, Arnaud Legout, and Nataliia Bielova. 2022. My Cookie is a phoenix: detection, measurement, and lawfulness of cookie respawning with browser fingerprinting. In *PETS 2022 - 22nd Privacy Enhancing Technologies Symposium*. Sydney, Australia. https://hal.science/hal-03218403 Accepted at the 22nd Privacy Enhancing Technologies Symposium (PETS 2022).

[40] Ari B. Friedman, Lujo Bauer, Rachel Gonzales, and Matthew S. McCoy. 2022. Prevalence of Third-Party Tracking on Abortion Clinic Web Pages. *JAMA Internal Medicine* 182, 11 (Nov 2022), 1221–1222. https://doi.org/10.1001/jamainternmed.2022.4208

[41] Alejandro Gómez-Boix, Pierre Laperdrix, and Benoit Baudry. 2018. Hiding in the Crowd: An Analysis of the Effectiveness of Browser Fingerprinting at Large Scale. In *Proceedings of the 2018 World Wide Web Conference* (Lyon, France) *(WWW '18)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 309–318. https://doi.org/10.1145/3178876.3186097

[42] Matthias Gotze, Srdjan Matic, Costas Iordanou, Georgios Smaragdakis, and Nikolaos Laoutaris. 2022. Measuring Web Cookies in Governmental Websites. In *14th ACM Web Science Conference 2022* (Barcelona, Spain) *(WebSci '22)*. Association for Computing Machinery, New York, NY, USA, 44–54. https://doi.org/10.1145/3501247.3531545

[43] Xiaodan Gu, Ming Yang, Congcong Shi, Zhen Ling, and Junzhou Luo. 2017. A novel attack to track users based on the behavior patterns. *Concurrency and Computation: Practice and Experience* 29, 6 (2017), e3891. https://doi.org/10.1002/cpe.3891 arXiv:https://onlinelibrary.wiley.com/doi/pdf/10.1002/cpe.3891 e3891 CPE-16-0089.R1.

[44] Hana Habib, Megan Li, Ellie Young, and Lorrie Cranor. 2022. "Okay, Whatever": An Evaluation of Cookie Consent Interfaces. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems* (New Orleans, LA, USA) *(CHI '22)*. Association for Computing Machinery, New York, NY, USA, Article 621, 27 pages. https://doi.org/10.1145/3491102.3501985

[45] Aniko Hannak, Gary Soeller, David Lazer, Alan Mislove, and Christo Wilson. 2014. Measuring Price Discrimination and Steering on E-Commerce Web Sites. In *Proceedings of the 2014 Conference on Internet Measurement Conference* (Vancouver, BC, Canada) *(IMC '14)*. Association for Computing Machinery, New York, NY, USA, 305–318. https://doi.org/10.1145/2663716.2663744

[46] Dominik Herrmann, Christian Banse, and Hannes Federrath. 2013. Behavior-Based Tracking: Exploiting Characteristic Patterns in DNS Traffic. *Comput. Secur.* 39 (nov 2013), 17–33. https://doi.org/10.1016/j.cose.2013.03.012

[47] Maximilian Hils, Daniel W. Woods, and Rainer Böhme. 2020. Measuring the Emergence of Consent Management on the Web. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 317–332. https://doi.org/10.1145/3419394.3423647

[48] Xuehui Hu, Guillermo Suarez de Tangil, and Nishanth Sastry. 2020. Multi-country Study of Third Party Trackers from Real Browser Histories. In *2020 IEEE European Symposium on Security and Privacy (EuroS&P)*. 70–86. https://doi.org/10.1109/EuroSP48549.2020.00013

[49] Xuehui Hu and Nishanth Sastry. 2020. What a Tangled Web We Weave: Understanding the Interconnectedness of the Third Party Cookie Ecosystem. In *12th ACM Conference on Web Science* (Southampton, United Kingdom) *(WebSci '20)*. Association for Computing Machinery, New York, NY, USA, 76–85. https://doi.org/10.1145/3394231.3397897

[50] Costas Iordanou, Georgios Smaragdakis, and Nikolaos Laoutaris. 2019. Who's Tracking Sensitive Domains? *CoRR* abs/1908.02261 (2019). arXiv:1908.02261 http://arxiv.org/abs/1908.02261

[51] Umar Iqbal, Steven Englehardt, and Zubair Shafiq. 2021. Fingerprinting the Fingerprinters: Learning to Detect Browser Fingerprinting Behaviors. In *2021 IEEE Symposium on Security and Privacy (SP)*. 1143–1161. https://doi.org/10.1109/SP40001.2021.00017

[52] Dongseok Jang, Ranjit Jhala, Sorin Lerner, and Hovav Shacham. 2010. An Empirical Study of Privacy-Violating Information Flows in JavaScript Web Applications. In *Proceedings of the 17th ACM Conference on Computer and Communications Security* (Chicago, Illinois, USA) *(CCS '10)*. Association for Computing Machinery, New York, NY, USA, 270–283. https://doi.org/10.1145/1866307.1866339

[53] Edden Kashi and Angeliki Zavou. 2020. Did I Agree to This? Silent Tracking Through Beacons. In *HCI for Cybersecurity, Privacy and Trust*, Abbas Moallem (Ed.). Springer International Publishing, Cham, 427–444.

[54] Matthias Kirchler, Dominik Herrmann, Jens Lindemann, and Marius Kloft. 2016. Tracked Without a Trace: Linking Sessions of Users by Unsupervised Learning of Patterns in Their DNS Traffic. In *Proceedings of the 2016 ACM Workshop on Artificial Intelligence and Security* (Vienna, Austria) *(AISec '16)*. Association for Computing Machinery, New York, NY, USA, 23–34. https://doi.org/10.1145/2996758.2996770

[55] Balachander Krishnamurthy and Craig Wills. 2009. Privacy Diffusion on the Web: A Longitudinal Perspective. In *Proceedings of the 18th International Conference on World Wide Web* (Madrid, Spain) *(WWW '09)*. Association for Computing Machinery, New York, NY, USA, 541–550. https://doi.org/10.1145/1526709.1526782

[56] Marek Kumpošt and Vašek Matyáš. 2009. User Profiling and Re-identification: Case of University-Wide Network Analysis. In *Trust, Privacy and Security in Digital Business*, Simone Fischer-Hübner, Costas Lambrinoudakis, and Günther Pernul (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 1–10.

[57] Anja Lambrecht and Catherine Tucker. 2019. Algorithmic Bias? An Empirical Study of Apparent Gender-Based Discrimination in the Display of STEM Career Ads. *Management Science* 65, 7 (2019), 2966–2981. https://doi.org/10.1287/mnsc.2018.3093 arXiv:https://doi.org/10.1287/mnsc.2018.3093

[58] Pierre Laperdrix, Benoit Baudry, and Vikas Mishra. 2017. FPRandom: Randomizing core browser objects to break advanced device fingerprinting techniques. In *Proc. of the Symposium on Engineering Secure Software and Systems (ESSOS)*. 97–114. https://hal.inria.fr/hal-01527580/file/fprandom-essos17.pdf

[59] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. 2015. Mitigating Browser Fingerprint Tracking: Multi-level Reconfiguration and Diversification. In *2015 IEEE/ACM 10th International Symposium on Software Engineering for Adaptive and Self-Managing Systems*. 98–108. https://doi.org/10.1109/SEAMS.2015.18

[60] Pierre Laperdrix, Walter Rudametkin, and Benoit Baudry. 2016. Beauty and the Beast: Diverting Modern Web Browsers to Build Unique Browser Fingerprints. In *2016 IEEE Symposium on Security and Privacy (SP)*. 878–894. https://doi.org/10.1109/SP.2016.57

[61] Pedro Leon, Ashwini Rao, Florian Schaub, Abigail Marsh, Lorrie Cranor, and Norman Sadeh. 2015. Privacy and Behavioral Advertising: Towards Meeting Users' Preferences. In *Eleventh Symposium On Usable Privacy and Security (SOUPS 2015)*. USENIX Association, Ottawa.

[62] Adam Lerner, Anna Kornfeld Simpson, Tadayoshi Kohno, and Franziska Roesner. 2016. Internet Jones and the Raiders of the Lost Trackers: An Archaeological Study of Web Tracking from 1996 to 2016. In *25th USENIX Security Symposium (USENIX Security 16)*. USENIX Association, Austin, TX. https://www.usenix.org/conference/usenixsecurity16/technical-sessions/presentation/lerner

[63] Song Li and Yinzhi Cao. 2020. Who Touched My Browser Fingerprint? A Large-Scale Measurement Study and Classification of Fingerprint Dynamics. In *Proceedings of the ACM Internet Measurement Conference* (Virtual Event, USA) *(IMC '20)*. Association for Computing Machinery, New York, NY, USA, 370–385. https://doi.org/10.1145/3419394.3423614

[64] Timothy Libert. 2015. Exposing the Hidden Web: An Analysis of Third-Party HTTP Requests on 1 Million Websites. *CoRR* abs/1511.00619 (2015). arXiv:1511.00619 http://arxiv.org/abs/1511.00619

[65] Bin Liu, Anmol Sheth, Udi Weinsberg, Jaideep Chandrashekar, and Ramesh Govindan. 2013. AdReveal: Improving Transparency into Online Targeted Advertising. In *Proceedings of the Twelfth ACM Workshop on Hot Topics in Networks* (College Park, Maryland) *(HotNets-XII)*. Association for Computing Machinery, New York, NY, USA, Article 12, 7 pages. https://doi.org/10.1145/2535771.2535783

[66] Elena Maris, Timothy Libert, and Jennifer R Henrichsen. 2020. Tracking sex: The implications of widespread sexual data leakage and tracking on porn websites. *New Media & Society* 22, 11 (2020), 2018–2038. https://doi.org/10.1177/1461444820924632 arXiv:https://doi.org/10.1177/1461444820924632

[67] David Martin, Hailin Wu, and Adil Alsaid. 2003. Hidden Surveillance by Web Sites: Web Bugs in Contemporary Use. *Commun. ACM* 46, 12 (dec 2003), 258–264. https://doi.org/10.1145/953460.953509

[68] Jonathan R. Mayer and John C. Mitchell. 2012. Third-Party Web Tracking: Policy and Technology. In *2012 IEEE Symposium on Security and Privacy*. 413–427. https://doi.org/10.1109/SP.2012.47

[69] Aleecia M. McDonald and Lorrie Cranor. 2011. A Survey of the Use of Adobe Flash Local Shared Objects to Respawn HTTP Cookies (CMU-CyLab-11-001). (1 2011). https://doi.org/10.1184/R1/6467741.v1

[70] Aleecia M. McDonald and Lorrie Faith Cranor. 2010. Americans' Attitudes about Internet Behavioral Advertising Practices. In *Proceedings of the 9th Annual ACM Workshop on Privacy in the Electronic Society* (Chicago, Illinois, USA) *(WPES '10)*. Association for Computing Machinery, New York, NY, USA, 63–72. https://doi.org/10.1145/1866919.1866929

[71] William Melicher, Mahmood Sharif, Joshua Tan, Lujo Bauer, Mihai Christodorescu, and Pedro Giovanni Leon. 2016. (Do Not) Track Me Sometimes: Users' Contextual Preferences for Web Tracking. *Proceedings on Privacy Enhancing Technologies* 2016 (2016), 135 – 154.

[72] Jakub Mikians, László Gyarmati, Vijay Erramilli, and Nikolaos Laoutaris. 2012. Detecting Price and Search Discrimination on the Internet. In *Proceedings of the 11th ACM Workshop on Hot Topics in Networks* (Redmond, Washington) *(HotNets-XI)*. Association for Computing Machinery, New York, NY, USA, 79–84. https://doi.org/10.1145/2390231.2390245

[73] Vikas Mishra, Pierre Laperdrix, Antoine Vastel, Walter Rudametkin, Romain Rouvoy, and Martin Lopatka. 2020. Don't Count Me Out: On the Relevance of IP Address in the Tracking Ecosystem. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 808–815. https://doi.org/10.1145/3366423.3380161

[74] Keaton Mowery, Dillon Bogenreif, Scott Yilek, and Hovav Shacham. 2011. Fingerprinting Information in JavaScript Implementations. In *Proceedings of W2SP 2011*. IEEE Computer Society.

[75] Keaton Mowery and Hovav Shacham. 2012. Pixel Perfect: Fingerprinting Canvas in HTML5. In *Proceedings of W2SP 2012*, Matt Fredrikson (Ed.). IEEE Computer Society.

[76] Mozilla. 2019. Firefox Now Available with Enhanced Tracking Protection by Default Plus Updates to Facebook Container Firefox Monitor and Lockwise. https://blog.mozilla.org/blog/2019/06/04/firefox-now-available-with-enhanced-tracking-protection-by-default/

[77] Martin Mulazzani, Philipp Reschl, Markus Huber, Manuel Leithner, Sebastian Schrittwieser, and Edgar R. Weippl. 2013. Fast and Reliable Browser Identification with JavaScript Engine Fingerprinting. In *Web 2.0 Security & Privacy (W2SP) 2013*.

[78] Shusuke Murai. 2016. Japan sticks with Internet Explorer as Microsoft ends support for old versions. Japan Times. https://www.japantimes.co.jp/news/2016/01/12/business/tech/japan-sticks-internet-explorer-microsoft-ends-support-old-versions/.

[79] Gabi Nakibly, Gilad Shelef, and Shiran Yudilevich. 2015. Hardware Fingerprinting Using HTML5. arXiv:1503.01408 [cs.CR]

[80] Joshua D. Niforatos, Alexander R. Zheutlin, and Jeremy B. Sussman. 2021. Prevalence of Third-Party Data Tracking by US Hospital Websites. *JAMA Network Open* 4, 9 (Sep 2021), e2126121–e2126121. https://doi.org/10.1001/jamanetworkopen.2021.26121

[81] Nick Nikiforakis, Wouter Joosen, and Benjamin Livshits. 2015. PriVaricator: Deceiving Fingerprinters with Little White Lies. In *Proceedings of the 24th International Conference on World Wide Web* (Florence, Italy) *(WWW '15)*. International World Wide Web Conferences Steering Committee, Republic and Canton of Geneva, CHE, 820–830. https://doi.org/10.1145/2736277.2741090

[82] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. 2013. Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. In *2013 IEEE Symposium on Security and Privacy*. 541–555. https://doi.org/10.1109/SP.2013.43

[83] Hartmut Obendorf, Harald Weinreich, Eelco Herder, and Matthias Mayer. 2007. Web Page Revisitation Revisited: Implications of a Long-Term Click-Stream Study of Browser Usage. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '07)*. Association for Computing Machinery, New York, NY, USA, 597–606. https://doi.org/10.1145/1240624.1240719

[84] Lukasz Olejnik, Claude Castelluccia, and Artur Janc. 2012. Why Johnny Can't Browse in Peace: On the Uniqueness of Web Browsing History Patterns. In *5th Workshop on Hot Topics in Privacy Enhancing Technologies (HotPETs 2012)*. Vigo, Spain. https://hal.inria.fr/hal-00747841

[85] Balaji Padmanabhan and Yinghui Yang. 2006. Clickprints on the Web: Are There Signatures in Web Browsing Data? *Information Technology & Systems eJournal* (10 2006). https://doi.org/10.2139/ssrn.931057

[86] Emmanouil Papadogiannakis, Panagiotis Papadopoulos, Nicolas Kourtellis, and Evangelos P. Markatos. 2021. User Tracking in the Post-Cookie Era: How Websites Bypass GDPR Consent to Track Users. In *Proceedings of the Web Conference 2021* (Ljubljana, Slovenia) *(WWW '21)*. Association for Computing Machinery, New York, NY, USA, 2130–2141. https://doi.org/10.1145/3442381.3450056

[87] Gaston Pugliese, Christian Riess, Freya Gassmann, and Zinaida Benenson. 2020. Long-Term Observation on Browser Fingerprinting: Users' Trackability and Perspective. *Proceedings on Privacy Enhancing Technologies* 2020 (2020), 558 – 577.

[88] Franziska Roesner, Tadayoshi Kohno, and David Wetherall. 2012. Detecting and Defending against Third-Party Tracking on the Web. In *Proceedings of the 9th USENIX Conference on Networked Systems Design and Implementation* (San Jose, CA) *(NSDI'12)*. USENIX Association, USA, 12.

[89] Jukka Ruohonen and Ville Leppänen. 2018. Invisible Pixels Are Dead, Long Live Invisible Pixels!. In *Proceedings of the 2018 Workshop on Privacy in the Electronic Society* (Toronto, Canada) *(WPES'18)*. Association for Computing Machinery, New York, NY, USA, 28–32. https://doi.org/10.1145/3267323.3268950

[90] Nayanamana Samarasinghe, Aashish Adhikari, Mohammad Mannan, and Amr Youssef. 2022. Et Tu, Brute? Privacy Analysis of Government Websites and Mobile Apps. In *Proceedings of the ACM Web Conference 2022* (Virtual Event, Lyon, France) *(WWW '22)*. Association for Computing Machinery, New York, NY, USA, 564–575. https://doi.org/10.1145/3485447.3512223

[91] Sebastian Schelter and Jérôme Kunegis. 2016. On the Ubiquity of Web Tracking: Insights from a Billion-Page Web Crawl. *CoRR* abs/1607.07403 (2016). arXiv:1607.07403 http://arxiv.org/abs/1607.07403

[92] Jonathan Schmidt. 2020. Does the dark side still have (ever)cookies? *FAU Erlangen-Nürnberg* (2020). https://faui1-files.cs.fau.de/public/publications/df/df-whitepaper-18.pdf

[93] Konstantinos Solomos, Panagiotis Ilia, Sotiris Ioannidis, and Nicolas Kourtellis. 2019. Clash of the Trackers: Measuring the Evolution of the Online Tracking Ecosystem. *CoRR* abs/1907.12860 (2019). arXiv:1907.12860 http://arxiv.org/abs/1907.12860

[94] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Hoofnagle. 2009. Flash Cookies and Privacy. *SSRN Electronic Journal* (08 2009). https://doi.org/10.2139/ssrn.1446862

[95] Statistia. 2022. https://www.statista.com/topics/1176/online-advertising/

[96] Elizabeth Stoycheff. 2023. Cookies and content moderation: affective chilling effects of internet surveillance and censorship. *Journal of Information Technology & Politics* 20, 2 (2023), 113–124. https://doi.org/10.1080/19331681.2022.2063215 arXiv:https://doi.org/10.1080/19331681.2022.2063215

[97] Latanya Sweeney. 2013. Discrimination in Online Ad Delivery: Google Ads, Black Names and White Names, Racial Discrimination, and Click Advertising. *Queue* 11, 3 (mar 2013), 10–29. https://doi.org/10.1145/2460276.2460278

[98] The Tor Project. 2023. Tor Browser—Tor Project Official Website. https://www.torproject.org/projects/torbrowser.html

[99] Ariana Tobin and Jeremy B. Merrill. 2018. Facebook Is Letting Job Advertisers Target Only Men. https://www.propublica.org/article/facebook-is-letting-job-advertisers-target-only-men

[100] Christof Ferreira Torres, Hugo Jonker, and Sjouke Mauw. 2015. FP-Block: Usable Web Privacy by Controlling Browser Fingerprinting. In *Computer Security – ESORICS 2015*, Günther Pernul, Peter Y A Ryan, and Edgar Weippl (Eds.). Springer International Publishing, Cham, 3–19.

[101] Erik Trickel, Oleksii Starov, Alexandros Kapravelos, Nick Nikiforakis, and Adam Doupé. 2019. Everyone is Different: Client-side Diversification for Defending Against Extension Fingerprinting. In *28th USENIX Security Symposium (USENIX Security 19)*. USENIX Association, Santa Clara, CA, 1679–1696. https://www.usenix.org/conference/usenixsecurity19/presentation/trickel

[102] United States Congress. 1986. Electronic Communication Privacy Act. https://www.congress.gov/bill/99th-congress/house-bill/4952 18 U.S.C. §§ 2510-2523.

[103] United States Supreme Court. 1976. United States v. Miller. https://supreme.justia.com/cases/federal/us/425/435/ 425 U.S. 435.

[104] United States Supreme Court. 1979. Smith v. Maryland. https://supreme.justia.com/cases/federal/us/442/735/ 442 U.S. 735.

[105] Blase Ur, Pedro Giovanni Leon, Lorrie Faith Cranor, Richard Shay, and Yang Wang. 2012. Smart, Useful, Scary, Creepy: Perceptions of Online Behavioral Advertising. In *Proceedings of the Eighth Symposium on Usable Privacy and Security* (Washington, D.C.) *(SOUPS '12)*. Association for Computing Machinery, New York, NY, USA, Article 4, 15 pages. https://doi.org/10.1145/2335356.2335362

[106] Tobias Urban, Martin Degeling, Thorsten Holz, and Norbert Pohlmann. 2020. Beyond the Front Page: Measuring Third Party Dynamics in the Field. In *Proceedings of The Web Conference 2020* (Taipei, Taiwan) *(WWW '20)*. Association for Computing Machinery, New York, NY, USA, 1275–1286. https://doi.org/10.1145/3366423.3380203

[107] Christine Utz, Martin Degeling, Sascha Fahl, Florian Schaub, and Thorsten Holz. 2019. (Un)Informed Consent: Studying GDPR Consent Notices in the Field. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 973–990. https://doi.org/10.1145/3319535.3354212

[108] Pelayo Vallina, Álvaro Feal, Julien Gamba, Narseo Vallina-Rodriguez, and Antonio Fernández Anta. 2019. Tales from the Porn: A Comprehensive Privacy Analysis of the Web Porn Ecosystem. In *Proceedings of the Internet Measurement Conference* (Amsterdam, Netherlands) *(IMC '19)*. Association for Computing Machinery, New York, NY, USA, 245–258. https://doi.org/10.1145/3355369.3355583

[109] Luca Vassio, Danilo Giordano, Martino Trevisan, Marco Mellia, and Ana Paula Couto da Silva. 2017. Users' Fingerprinting Techniques from TCP Traffic. In *Proceedings of the Workshop on Big Data Analytics and Machine Learning for Data Communication Networks* (Los Angeles, CA, USA) *(BigDAMA '17)*. Association for Computing Machinery, New York, NY, USA, 49–54. https://doi.org/10.1145/3098593.3098602

[110] Antoine Vastel, Pierre Laperdrix, Walter Rudametkin, and Romain Rouvoy. 2018. Fp-Scanner: The Privacy Implications of Browser Fingerprint Inconsistencies. In *27th USENIX Security Symposium (USENIX Security 18)*. USENIX Association, Baltimore, MD, 135–150. https://www.usenix.org/conference/usenixsecurity18/presentation/vastel

[111] Antoine Vastel, Pierre Laperdrix, Walter Rudametkin, and Romain Rouvoy. 2018. FP-STALKER: Tracking Browser Fingerprint Evolutions. In *2018 IEEE Symposium on Security and Privacy (SP)*. 728–741. https://doi.org/10.1109/SP.2018.00008
[112] Webkit.org. 2025. Intelligent Tracking Prevention. https://webkit.org/blog/7675/intelligent-tracking-prevention/
[113] Ben Weinshel, Miranda Wei, Mainack Mondal, Euirim Choi, Shawn Shan, Claire Dolin, Michelle L. Mazurek, and Blase Ur. 2019. Oh, the Places You've Been! User Reactions to Longitudinal Transparency About Third-Party Web Tracking and Inferencing. In *Proceedings of the 2019 ACM SIGSAC Conference on Computer and Communications Security* (London, United Kingdom) *(CCS '19)*. Association for Computing Machinery, New York, NY, USA, 149–166. https://doi.org/10.1145/3319535.3363200
[114] Vera Wesselkamp, Imane Fouad, Cristiana Santos, Yanis Boussad, Nataliia Bielova, and Arnaud Legout. 2021. In-Depth Technical and Legal Analysis of Tracking on Health Related Websites with ERNIE Extension. In *Proceedings of the 20th Workshop on Workshop on Privacy in the Electronic Society* (Virtual Event, Republic of Korea) *(WPES '21)*. Association for Computing Machinery, New York, NY, USA, 151–166. https://doi.org/10.1145/3463676.3485603
[115] Andrew G. West and Adam J. Aviv. 2014. Measuring Privacy Disclosures in URL Query Strings. *IEEE Internet Computing* 18, 6 (2014), 52–59. https://doi.org/10.1109/MIC.2014.104
[116] Craig E. Wills and Can Tatar. 2012. Understanding What They Do with What They Know. In *Proceedings of the 2012 ACM Workshop on Privacy in the Electronic Society* (Raleigh, North Carolina, USA) *(WPES '12)*. Association for Computing Machinery, New York, NY, USA, 13–18. https://doi.org/10.1145/2381966.2381969
[117] Shujiang Wu, Song Li, Yinzhi Cao, and Ningfei Wang. 2019. Rendered Private: Making GLSL Execution Uniform to Prevent WebGL-based Browser Fingerprinting. In *28th USENIX Security Symposium (USENIX Security 19)*. USENIX Association, Santa Clara, CA, 1645–1660. https://www.usenix.org/conference/usenixsecurity19/presentation/wu
[118] Yinghui (Catherine) Yang. 2010. Web user behavioral profiling for user identification. *Decision Support Systems* 49, 3 (2010), 261–271. https://doi.org/10.1016/j.dss.2010.03.001
[119] Zhiju Yang and Chuan Yue. 2020. A Comparative Measurement Study of Web Tracking on Mobile and Desktop Environments. *Proceedings on Privacy Enhancing Technologies* 2020 (Apr 2020), 24–44. https://doi.org/10.2478/popets-2020-0016
[120] Xiufen Yu, Nayanamana Samarasinghe, Mohammad Mannan, and Amr Youssef. 2022. Got Sick and Tracked: Privacy Analysis of Hospital Websites. In *2022 IEEE European Symposium on Security and Privacy Workshops (EuroS&PW)*. 278–286. https://doi.org/10.1109/EuroSPW55150.2022.00034

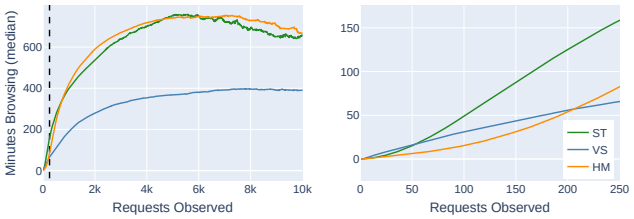


Figure 11: Average requests observed over the amount of time since the user first started browsing.

A Appendix

A.1 Observed requests over time

In Section 4.4.1, we examined how users lose anonymity as they browse more and more web pages which in turn allows the tracker to gain a greater amount of information about them. To add additional context to these results, Figure 11 shows the median time, in minutes, between a user starting browsing and a tracker observing a given number of requests. The left subplot shows the full range from 0 to 10,000 requests while the right subplot shows the same curves but from 0 to 250 requests (indicated by the dotted line in the left subplot). Based on this information, 5 requests are typically made within the first 15–60 seconds of being online. 30 requests are usually made within the first 3.5–10 minutes and 100–250 requests anywhere between 15–50 and 65–160 minutes depending on the dataset. That means users on average lose 78–85% of their anonymity within the first 60 seconds and 90% anonymity within the first 10 minutes of browsing the web.