Justin Brookman, Phoebe Rouge, Aaron Alva, and Christina Yeung

Cross-Device Tracking: Measurement and Disclosures

Abstract: Internet advertising and analytics technology companies are increasingly trying to find ways to link behavior across the various devices consumers own. This *cross-device tracking* can provide a more complete view into a consumer's behavior and can be valuable for a range of purposes, including ad targeting, research, and conversion attribution. However, consumers may not be aware of how and how often their behavior is tracked across different devices. We designed this study to try to assess what information about cross-device tracking (including data flows and policy disclosures) is observable from the perspective of the end user. Our paper demonstrates how data that is routinely collected and shared online could be used by online third parties to track consumers across devices.

Keywords: privacy, cross-device, tracking, transparency

DOI 10.1515/popets-2017-0020
Received 2016-08-31; revised 2016-11-30; accepted 2016-12-01.

1 Introduction

Concerns about cross-device tracking often center on the unknown: who is performing this tracking, when does it occur, and how is it performed? To help shine some light on this topic, we conducted a review of 100 popular websites to determine which sites transmit data or otherwise perform actions known to facilitate cross-device tracking. We observed the following data collection practices that could be used by tracking companies to make inferences about linkages among devices:

– We detected connections to the same third party services on different devices. When those devices share common attributes — such as the same local network and IP address — those services may be able to correlate user activity across devices. In visiting 100 sites on two virtual devices, we connected to 861 different third party domains on both devices, including domains operated by dedicated cross-device tracking companies.
– 96 out of 100 of the sites we tested allowed consumers to submit a username or email address that could be shared to correlate users across devices.
– Six companies that collect login information as a first party also collect extensive behavioral data as a third party on other websites.
– 16 out of 100 sites shared personally identifying information with third parties, either in raw or hashed form.
– Dozens of third party services sync unique cookie ID values, which could facilitate the sharing of cross-device graph information.

These findings demonstrate that a broad range of companies possess the capacity to correlate user behavior across different devices that the users own. However, although the data practices summarized above could be used for cross-device linkage, we could not definitively determine that the data was used by a company for that purpose in any particular instance. From the perspective of the consumer, it is challenging to know when cross-device tracking occurs, since companies can make determinations of device correlation on their own servers, unobservable to end users (we did not detect companies using the same cookie IDs across devices to identify linked devices). The data transmissions observed could be for purposes other than cross-device tracking, though a user may be unable to rule out the possibility based on the data alone.

Because the purposes for which the data transmitted was often ambiguous, we also looked at the privacy disclosures of these 100 websites in order to see if they made information available about cross-device tracking. Most of the policies we reviewed reserve broad rights to allow third parties to collect and use pseudonymous browser data such as IP address and unique cookie identifiers. However, the website privacy policies we reviewed contained little explicit discussion of cross-device

Justin Brookman: Federal Trade Commission, Office of Technology Research and Investigation, E-mail: jbrookman@ftc.gov
Phoebe Rouge: Federal Trade Commission, Office of Technology Research and Investigation, E-mail: prouge@ftc.gov
Aaron Alva: Federal Trade Commission, Office of Technology Research and Investigation, E-mail: aalva@ftc.gov
Christina Yeung: Federal Trade Commission, Office of Technology Research and Investigation, E-mail: cyeung@ftc.gov

tracking specifically, or whether consumers had the ability to turn off cross-device linkages.

2 Background

Since the late 1990s, the Federal Trade Commission (FTC)[1] has sought to bring greater transparency and user control to the issue of online behavioral data collection as part of its work to protect and promote consumer privacy.[2] In 2007, Commission staff held a two-day workshop[3] focused specifically on behavioral targeting. In 2009, staff published "Self-Regulatory Principles for Online Advertising" [2] to encourage the adoption of more transparent practices. Behavioral advertising was also a significant focus of the 2012 Report "Protecting Consumer Privacy in an Era of Rapid Change: Recommendations for Businesses and Policymakers" [3]. The Commission has also brought numerous enforcement actions to stop unfair and deceptive practice related to online behavioral advertising [4–6].

Despite the progress to date addressing these issues, consumers continue to have concerns about online privacy and tracking. According to a 2016 TRUSTe survey, [7] 92% of consumers worry about their privacy online. With regard to online behavioral data collection, a recent Pew survey [8] shows that over three quarters of internet users are not confident that online advertisers will maintain the privacy and security of their web browsing data. And, while most expressed an interest in controlling the collection of their online data, few felt empowered to do so [8].

Originally, online behavioral data collection was limited to connecting users across multiples websites on one device. Today, advertising technology companies are finding ways to track users across devices as well. Consumers interact with more devices — and smarter devices — than ever before, including computers, smartphones, smart TVs and Blu-Ray players, gaming platforms, and Internet of Things devices. Connecting this data can be valuable to companies not just for ad targeting, but also for research, security, and purchase attribution purposes (e.g., if an ad on one device resulted in a sale on another). The commercial benefits from cross-device tracking could potentially be passed on to consumers in the form of lower prices, though we have not analyzed that issue here. This paper does not seek to weigh the potential benefits of cross-device tracking with any privacy risks or interests; instead, it explores what information concerning cross-device tracking is discoverable by consumers — through disclosures made to consumers in privacy policies, and through observable behavior of websites in browsers.

Staff of the Office of Technology Research and Investigation conducted this research in conjunction with the FTC staff's November 16, 2015 workshop on Cross-Device Tracking. At that event, stakeholders from industry, academia, and civil society gathered to explore the unique privacy issues associated with the practice.[4]

This paper will initially describe some of the current models that companies use to compile "device graphs" for consumers — that is, lists of device identifiers that the company imputes to a particular individual. We then describe the methodology for the study we ran on 100 popular websites to observe the collection of data that could be used for cross-device tracking purposes. We then present the full findings of the study which were briefly summarized in the abstract: we describe the numerous cases in which data is transmitted that could be used to facilitate cross-device tracking. Next, we describe the results of our review of the privacy policies of those 100 websites. Finally, we discuss the limitations of our study, and conclude with suggestions for possible future research.

While several previous studies [9–12] have considered data collection online, including third-party data collection, we are unaware of any prior studies that considered such collection in the context of cross-device tracking across such a large number of sites. Other work [13–16] has focused on third-party data collection across mobile applicatons (as opposed to web browsers); while we do not measure leakage of information through apps, many of the same privacy interests are implicated. A number of papers [17, 18] have also looked specifically at the efficacy and completeness of disclosures in privacy policies though, again, without regard to cross-device tracking. Finally, some researchers have looked at non-cookies tracking mechanisms such as Flash Cookies [19]

1 This research has been prepared by staff members of the Office of Technology Research and Investigation of the Federal Trade Commission. The views expressed do not necessarily reflect the views of the Commission or any individual Commissioner.

2 *See* https://www.ftc.gov/news-events/events-calendar/1999/11/online-profiling-public-workshop and [1]

3 https://www.ftc.gov/news-events/events-calendar/2007/11/ehavioral-advertising-tracking-targeting-technology

4 https://www.ftc.gov/news-events/events-calendar/2015/11/cross-device-tracking

or digital fingerprinting [21]. Our paper focuses primarily on how often sites connect to third-party services; we do not analyze in detail what methods companies might use to keep state on user activity. We do look at how *cookie syncing* could be used to share device graphs among third-party services building off of previous research [20] into methods of detecting such syncing.

3 Description of Cross-Device Tracking Models

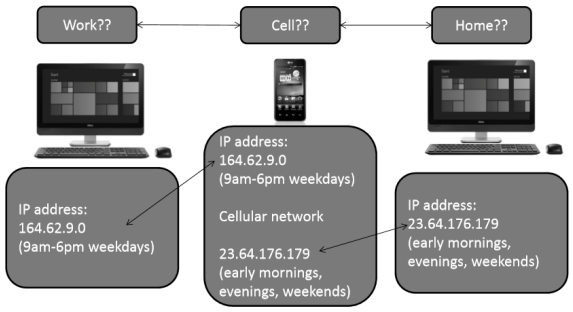
Below we summarize some of the primary models that are in use today to link behavior across multiple devices. This discussion presumes a basic understanding of how online data collection operates, including the collection of IP addresses and the placing and reading of HTML cookies by first and third parties [23].

3.1 Probabilistic Cross-Device Tracking

Probabilistic cross-device tracking works by first uniquely identifying various devices (for example, through a cookie, hardware identifier, or device fingerprint[5]), and then comparing collected information about those devices for shared attributes to infer a likelihood of whether those devices are used by the same person. IP address is one such identifier: If two devices repeatedly use the same IP address — because they are both using the same wireless router to connect to the internet — there is a significant possibility that those two devices are owned or used by the same person. Connected devices routinely connect to the internet through shared networks, so comparing shared IP addresses is one of the easiest — if most rudimentary — tactics for assessing common ownership of devices (see Figure 1).

In the example below, a smartphone shares an IP address with one computer during weekdays and with a different computer on nights and weekends. From this information alone, a company's algorithm could assess a likelihood that the three devices are used by the same person (i.e., that the first computer and the phone share a common WiFi connection during business hours, and the second computer and the phone share a different WiFi connection during non-business hours). With this determination about the likelihood of the use, a company could link the devices on a device graph. If a company is also able to access geolocation information from each of the devices and determines a similar correlation of shared physical location it will have greater confidence that the devices are used by the same person.

Fig. 1. Probabilistic Matching



However, correlation of IP address and location does not necessarily guarantee that the same person is using the devices. For instance, several devices owned by different customers may share a WiFi connection over a coffee shop's WiFi network at any point in time. Thus, probabilistic companies may try to factor in other signals as well to further increase their confidence in their models [22]. For example, if a company notices similar browsing patterns across certain devices (for example, the user of each devices regularly visits a New Orleans Pelicans fan site, a certain technology blog, and a Capitol Hill community site), the company may be able to ascribe device correlation with more certainty, using proprietary algorithms that determine that the similar browsing behavior across the devices is likely being performed by the same user [26]. Estimates on the accuracy of probabilistic device correlations range as high as 97.3% [27]. That is, even if users never share identifiers such as an email address or user name, companies that use probabilistic device tracking may be able to correctly link devices over 97% of the time.

5 For an overview of device fingerprinting, see [24]. As web browsers become more sophisticated and intricate, they may introduce new elements of entropy that can distinguish an individual browser to make it identifiable over time. For instance, some fingerprinting techniques rely on identifying subtle and barely perceptible differences in how browsers render certain images in order to fingerprint users [25].

3.2 Deterministic Cross-Device Tracking

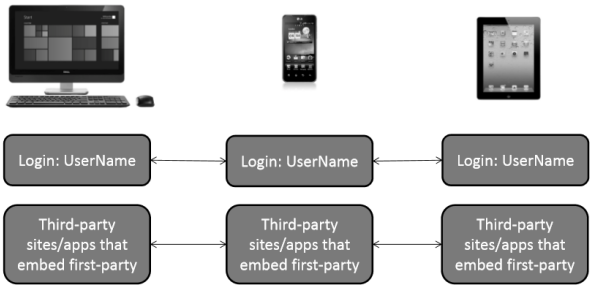
Deterministic cross-device tracking involves tying unique devices to a common persistent identifier — such as a name or email address. Although web browsing has been traditionally described as "anonymous," [28, 29] in many situations consumers affirmatively provide identifying information to various websites, such as when they create an account or login to a site. Leveraging this identifying information, companies may be able to correlate user activity across other devices where the consumer uses the same credentials. Companies may also share identifying information with third party data brokers that do not have a direct consumer relationship to allow those entities to engage in cross-device tracking.

3.2.1 Logged-in, cross-context tracking

Many services allow users to log into personal accounts from any internet-connected device — thus, a consumer can access her email or a social networking account from a desktop computer, her phone, or a friend's laptop. If accessed through a web browser, the service can place a cookie on the device to remember the user in the future. Other devices may allow similar tools to allow the service to recognize the device in the future, such as a device-specific identifier on smartphones. Keeping track of the specific devices that a consumer uses to access an account can be useful for security purposes. If there is an attempt to login to an account from a new device, there is a greater chance that that login may be fraudulent, and the service may want to require another level of authentication before granting access to the account. Many services will even show a consumer the devices that have logged into her account, so that she can monitor potentially unauthorized access.

Many sites that offer login capability also offer functionality that can be embedded into other sites as well — such as social sharing widgets, analytics code, social login, or advertising. If a user is logged into a service that provides embedded functionality to another website, the service may log the fact that the user visited that site — regardless of whether the user interacts with the embedded content. This consumer's viewing history is then part of the service's profile of the consumer. If a consumer logs into a service account on different devices, such as a work computer, phone, and a home tablet, that service may have the ability to track the consumer across a wide range of web browsing on all three devices, which may not be apparent to the consumer (see Figure 2).

Fig. 2. Logged-in deterministic matching



3.2.2 Shared credential cross-device tracking

Other third-party tracking companies do not have login relationships with consumers directly, but may have contractual relationships with other companies that do. For example, sites could pass along identifying information during login to a tracking company, [12] allowing that tracking company to match user profiles on other devices. This information passed along could be the login credential itself (say, an email address), or it could be a cryptographic hash of the identifier.[6] If a hash is transmitted to the same third party from different devices as in Figure 3, the company could match the strings together across hashes of the same identifiers received on

6 A hash is a one-way mathematical function that turns any amount of data into a different fixed length value. The hash value has no clear connection to the input, so it is difficult, given only the hashed output, to reverse the hash to find the original value. However, every time a hash of any single input is calculated, the output is the same. Hashing algorithms (such as MD5 and SHA-256) are designed to avoid "collision" — that is, two different inputs resulting in the same output. Thus, if two hashed outputs match, it is highly likely that the inputs were the same as well.
As an example, the MD5 hash of the identifier "user@domain.com" will always be "cd2bfcffe5fee4a1149d101994d0987f."
For more discussion on hashing — as well as the practical limitation of hashes — see [30].

Fig. 3. Matching across publishers using hashed identifiers

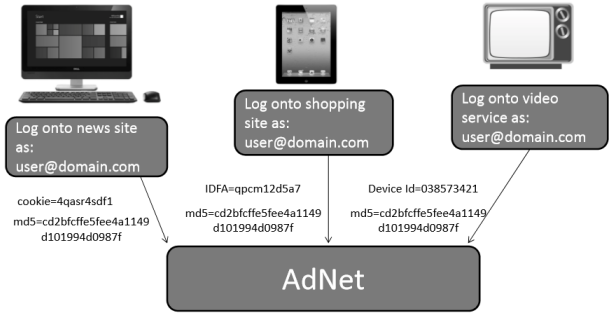
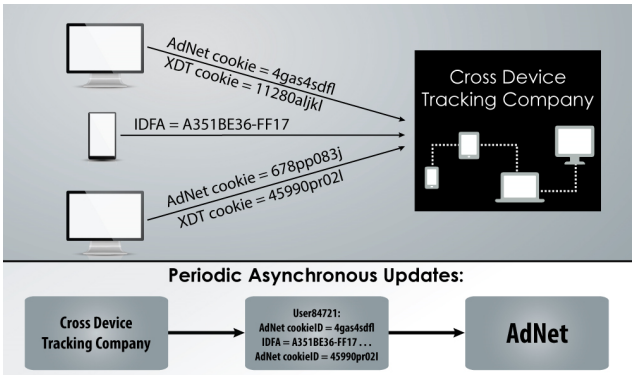


Fig. 4. Graph Sharing



other devices without ever collecting the actual login identifier itself.

In this example, by having a relationship with sites that collect login credentials (or otherwise collect an identifier such as email address), the third party advertising network is sent hashes of that identifier on three different devices, and is thus able to track the user across the three devices.

3.2.3 Other cross-device tracking models

The techniques described above are deliberately simplistic for the purpose of explanation; in practice, companies may use different methods altogether, or engage in a combination of the above techniques. Some companies may simply purchase or lease a cross-device graph from other companies. One method for sharing device graphs is for one company to send its identifiers (such as cookie values) to a cross-device tracking company via cookie syncing; [20] the cross-device tracking company could then transmit back to the original company a graph of devices it has identified as linked to the same user (see Figure 4).

Some companies blend probabilistic and deterministic methods in various ways. For example, a cross-device tracking company that relies primarily on probabilistic methods might contract with a deterministic company to verify that its matching algorithms are accurate. Alternatively, a primarily deterministic company might expand its data set to include likely probabilistic matches for clients who prefer broader reach to certainty.

And companies are constantly developing novel methods to correlate users across devices. For example, some are embedding unique ultrasonic frequencies into television content that that may be detected by an always-on microphone on a smartphone [31]. However, those and other novel methods of cross-device tracking are outside the scope of this study.

4 Study

For our study, we used two different virtual devices to browse the same 100 popular websites — the top 20 sites for Games, Sports, News, Shopping, and Reference according to the web metrics company Alexa.[7] On each virtual device, we conducted two complete runs of the same 100 sites. When browsing each site, we identified the third parties that received data and monitored the specific data they received. The goal of the study was to observe data collection across both devices that could be used to facilitate the cross-device tracking techniques described above, and to detect whether companies use the same client-side identifiers across multiple devices associated with the same user.

4.1 Data Collection Methodology

We used OpenWPM — an open source web privacy measurement platform developed at Princeton University — as the platform to facilitate our data collection.[8] We used OpenWPM to collect all data and to automate

7 As of October 15, 2015.

8 https://github.com/citp/OpenWPM

portions of the navigation to each of the primary domains we visited. To set up the devices, we first created one Ubuntu virtual device, or virtual machine (VM1). We installed all Ubuntu security updates, OpenWPM dependencies, and the OpenWPM tool itself. Then we cloned the virtual machine to make a second identical machine (VM2). We modified the MAC address on VM2 and connected it externally using the same local IP address in order to appear to web domains as a separate computer on the same local network.

We began the study by using VM1 to create free email and social media accounts with Google, Yahoo, Facebook, and Microsoft Outlook. Because users often log into these services across multiple devices, information associated with these accounts could potentially be used to correlate user behavior across devices. These accounts each referred to the same (fictitious) person, and used the same username.

We then proceeded to visit our 100 sites on the two devices four different times. For each of our four runs of the 100 sites, we initiated the OpenWPM instance, which automatically navigated to the first domain. When we navigated to a domain, we allowed the homepage to load completely, then manually clicked on a prominent link. For news and sports sites, this link was a story on the homepage. For other sites, it was the most prominent non-advertising link/banner that we could find. At the end of both phases, OpenWPM saved the browsing data, and also saved a profile of the browser.

We completed two of the four runs on VM1 (Run 1 and Run 3) and two runs on VM2 (Run 2 and Run 4). The goal of the experiment was to simulate the experience of a user who browses similar sites on different devices, and to detect when information is shared with third parties that could be used to correlate users across devices. For Run 1, in addition to the email and social media profiles described above, we used VM1 to create an account on each site where account creation was possible. This is likely more logins than a typical user would perform (consumers routinely visit sites without creating accounts), but we wanted to test all opportunities where identifying login information could be conveyed to third parties. When a site required an email address to login or sign up, we provided the same Gmail address to each site. For Run 2 — using VM2 — we began by logging in to the email and social media profiles we first created on VM1. We also checked our Gmail accounts for any emails that requested us to confirm our newly created accounts. For each new account that requested clicking a confirmation link, we did so. Then we proceeded to login to each account possible as we proceeded sequentially through the test of 100 sites. For each login, we used the credentials we had set up on VM1 during Run 1. By taking advantage of every opportunity to provide identifying information to first-party sites on both devices, we maximized our ability to detect when identifying information might be shared for deterministic cross-device tracking.

During Run 3 (on VM1) and Run 4 (on VM2) we navigated the browser as a typical consumer without affirmatively logging onto any service (either the email or social media profiles, or the 100 test sites). In many cases, however, we observed that we were already logged into accounts when we revisited our test sites that we had logged into on Run 1 and Run 2 — much like a browser would with a consumer that has reopened her browser without clearing browsing data. Running these two additional tests on both devices gave us more robust data from which to draw conclusions about third-party data sharing.

4.2 Data Analysis

Once we had collected our data, we reviewed the data to find evidence of information sharing that could be used for third-party cross-device tracking. Specifically, we looked for (1) instances of connections to third-party domains and (2) instances of sharing identifiers or hashes of identifiers with third parties.

The output for each run was a SQLite database that was structured, formatted, and produced according to OpenWPM's specifications [32]. We parsed out the data contained in the database to isolate features of the URL, such as hostname, path, and query string, as well as cataloged all the different key-value pairs found in the query string and parameters. We also attempted to classify HTTP requests and responses as first party or third party via simple string matching of the hostname in an individual URL to the hostname of the site being visited in the browser. In the final analysis, this was supplemented by visual inspection to identify false positives (e.g., separate domains, such as a content display network, that appeared to be operated by the first party) in some cases. For performance reasons, we imported the combined database into MS SQL Server 2012 to run the analysis.

We then used a combination of SQL queries and Python scripts to search the data and presented it in spreadsheet form. We looked specifically for evidence of the same identifiers being shared with third-parties dur-

ing different browsing runs, and on the different VMs. We also generally examined the composition of data collection by third-parties over a combination of all four runs. We specifically searched for evidence of HTTP requests that included the personally identifiable information of the user account (e.g., email, full name, user IDs).[9] Finally, we adapted the code previously developed by Princeton researchers and available publically on Github,[10] to search for evidence of domains sharing device-specific cookie values with other domains (also known as "cookie syncing" (see infra, 5.4)). Although this code likely did not identify all instances of cookie syncing, it did highlight numerous examples of domains sharing high entropy cookie values (i.e., sufficiently distinct identifiers that could uniquely identify a device) with other third-party domains.[11]

5 Results

Our study demonstrated extensive third-party data collection that could be used to enable cross-device tracking. Our primary findings are:

- 861 third-party domains collected data across test runs on both virtual devices that could be used for probabilistic device linkage. Companies that specialize in probabilistic cross-device linkage collected data on 34% of the sites we visited.
- Six large first parties who enable user login also collect extensive third-party data across multiple devices.
- At least 16 out of the 100 sites we reviewed shared personally identifiable information — or hashed personally identifiable information — with third parties, which could allow third parties to correlate multiple devices to persistent real world identifiers.
- 106 third-party domains shared unique, browser-specific cookie identifiers with 210 other third parties — including dedicated cross-device tracking companies — potentially enabling third party tracking companies to share device graph information with each other.

We did not detect third parties using the same cookie values across multiple devices, which would have shown conclusive instances of cross-devices tracking. However, the use of shared cookie values would not necessarily be expected even if cross-device tracking were occurring, as companies could store information about correlated identifiers on their own systems.

5.1 Numerous trackers see users across multiple websites on both devices

In our scan of 100 sites, we detected extensive communication to third-party domains on the considerable majority of sites. These third-party domains are predominantly operated by advertising and analytics companies; by virtue of being embedded (directly or indirectly)[12] within the first-party site, these companies can place or read a unique domain-specific cookie identifying the device, as well as the identity of the first party site the user visited. Table 1 provides an overview of the sites that connected to the most third-party services. Thirteen of the 100 sites connected to over 100 different domains on at least one of our four visits to the site; one site connected to an average of 122 third-party domains over the course of four different runs. The median site connected to an average of 37.5 third-party domains during our tests.

In visiting these 100 sites four times each, our test browser was connected to a total of 1130 additional distinct domains (see Table 2). for the third-party domains with the most connections). Some of these were other domains owned by the same first party — for example, the site NHL.com directed traffic to nhle.com, another domain that the NHL operates. Additionally, several of these third-party domains were operated by the same third party (e.g., Google oper-

9 We looked for this information in both plain, hashed, (MD5, SHA-1, SHA-256) and base64 encoded forms.

10 https://github.com/citp/TheWebNeverForgets

11 For example, the script identified the domain geo-um.btrll.com setting a cookie with an identifier BR_APS=3VjoOnCnch8EBKBheSg. That same identifier was sent to 46 different domains, and later appears in a tracking cookie composed of multiple identifiers from multiple domains set by one of these domains, sync.adaptv.advertising.com.

12 A first-party publisher might only code its site to direct to a handful of third-party advertising companies. However, once a browser has connected to one of those third parties, those companies may then initiate connections to any number of additional third parties. Indeed, this is how programmatic real-time bidding for advertising impressions often work: A publisher may embed code for a supply side platform or ad exchange on its site. That site will then connect to several other third parties to see which will bid the most to serve a particular ad impression to the consumer [33].

Table 1. Top 20 sites with the most connections to third-party domains on 100 sites tested

First Party Domain	Run1	Run2	Run 3	Run 4	Average
timesofindia.indiatimes.com	102	156	166	65	122.25
bbc.co.uk/news	108	81	107	88	96
weather.com	69	150	109	12	85
wowhead.com	106	63	73	90	83
pcgamer.com	67	83	81	94	81.25
bbc.co.uk/sport/o/football	95	112	63	44	78.5
time.com	97	86	72	57	78
goal.com	169	59	33	50	77.75
foxnews.com	109	78	86	38	77.75
huffingtonpost.com	121	77	51	56	76.25
usatoday.com	109	87	59	43	74.5
nbcnews.com	42	97	73	84	74
thesaurus.com	67	55	43	122	71.75
forbes.com	108	106	—[a]	72	71.75
sbnation.com	38	70	107	65	70
cbssports.com	48	67	82	72	67.25
reuters.com	76	78	54	60	67
walmart.com	38	40	95	84	64.25
nytimes.com	58	60	83	52	63.25
nhl.com	61	57	59	71	62

a Our test browser crashed during Run3 for Forbes resulting in no data collection for that run. For this reason, the average number of third-party connections for forbes.com is almost certainly artificially low.

Table 2. Top 20 third-party domains with most connections from 100 sites tested

Third-Party Domain	Run1	Run2	Run3	Run4	Average
doubleclick.net	88	89	87	86	87.5
facebook.com	69	71	68	68	69
google.com	70	69	70	62	67.75
google-analytics.com	65	67	64	58	63.5
scorecardresearch.com	65	60	61	58	61
googlesyndication.com	62	63	58	58	60.25
adnxs.com	48	47	48	50	48.25
2mdn.net	48	49	44	46	46.75
gstatic.com	49	55	4	34	46
googleapis.com	47	54	38	43	45.5
cloudfront.net	46	48	44	41	44.75
yahoo.com	47	50	44	36	44.25
moatads.com	47	46	42	40	43.75
bluekai.com	44	45	40	39	42
twitter.com	43	41	40	32	39
advertising.com	35	41	42	37	38.75
rubiconproject.com	40	37	38	39	38.5
adsafeprotected.com	38	38	41	35	38
rlcdn.com	34	38	38	37	36.75
imrworldwide.com	37	39	38	32	36.5

ates doubleclick.net, google.com, google-analytics.com, googleapis.com, 2mdn.net, and others). Nevertheless, the 100 most common third-party domains still represented 64 separate companies.[13]

Most relevant to cross-device tracking, 861 domains saw users on at least one run on each device. 612 domains saw users on all four runs of the same 100 websites. Any third party company with visibility into two different browsers would have the capacity to engage in at least rudimentary probabilistic cross-device tracking based on IP address alone (for devices that share IP address on a common network, such as a home router). As discussed supra, 3.1 wide deployment across a greater number of first-party sites would give third party companies a greater ability to develop behavioral profiles on each of the devices — as well as ascribe likelihood of common ownership based on similar behavioral profiles.[14] 432 third party domains saw users across more than one first-party site on both devices.

A number of the third party services publicly purport to specialize in probabilistic cross-device tracking.[15] Our test browsers were frequently directed to the servers of probabilistic cross-devices tracking companies on each of our four test runs of the 100 websites. On average, 34.25 of 100 sites connected to one or both of two leading probabilistic cross-device tracking services. As discussed below, not only can this exposure facili-

13 The identity of the operators of these domains was determined by searching the publicly available WHOIS database, as well as by referencing the databases operated by Ghostery and Cookiepedia. In some cases, we were unable to determine conclusively who operated a particular domain; 14 of the top 100 third-party domains were registered by privacy proxy services in the WHOIS database.

14 The number of connections discussed above only reflects connections that are made through the consumer's browser. Alternatively, a first or third party may contact a third party advertising or analytics service directly without involving the user's browser at all. These *server-to-server* communications about the consumer's device would be invisible to the consumer, unless the recipient of a server-to-server communication subsequently connected to the browser (e.g., because it had won an auction to serve an advertisement). Thus, the number of connections to third-party services may be substantially understated.

15 *See, e.g.*, http://www.tapad.com/about-us/who-we-are/ ("Tapad Inc. is a marketing technology firm renowned for its breakthrough, unified, cross-device solutions.") and https://drawbridge.com/c/graph ("The Drawbridge Connected Consumer Graph is the industry's leading cross-device identity solution, reaching more than one billion consumers across more than five billion digital touchpoints.") Numerous other companies likely make assumptions about connected devices through probabilistic methods such as IP address correlation, but these are the only two companies that looked to attempt to quantify potential probabilistic cross-device tracking.

Table 3. First party login sites with third-party reach

Third-Party Domain	Run1	Run2	Run3	Run4	Average
doubleclick.net (Google)	88	89	87	86	87.5
facebook.com	69	71	68	68	69
google.com	70	69	70	62	67.75
google-analytics.com	65	67	64	58	63.5
googlesyndication.com	62	63	58	58	60.25
2mdn.net (Google)	48	49	44	46	46.75
gstatic.com (Google)	49	55	46	34	46
googleapis.com	47	54	38	43	45.5
cloudfront.net (Amazon)	46	48	44	41	44.75
yahoo.com	47	50	44	36	44.25
twitter.com	43	41	40	32	39
advertising.com (Verizon)	35	41	42	37	38.75
facebook.net	40	39	30	26	33.75
googleadservices.com	32	36	27	21	29
amazon-adsystem.com	23	22	27	25	24.25
googletagmanager.com	22	25	22	25	23.5
adtechus.com (Verizon)	20	23	21	20	21
amazonaws.com	19	23	22	15	19.75
adap.tv (Verizon)	13	16	22	15	17.25
liverail.com (Facebook)	16	18	20	14	17
youtube.com (Google)	12	18	17	11	14.5
fbcdn.net (Facebook)	6	18	13	14	12.75
ytimg.com (Google)	10	14	17	10	12.75
yimg.com (Yahoo)	11	15	12	11	12.25
atdmt.com (Facebook)	10	10	11	13	11
atwola.com (Verizon)	10	10	11	11	10.5
convertro (Verizon)	13	8	9	8	9.5

tate those companies' cross-device tracking of a user, but through cookie-syncing, can allow other companies to purchase the cross-device graphs that they — and other similar companies — can develop.

5.2 Logged-in Deterministic Tracking

As described, supra 3.2.1, several sites that enable users to log in as a first party also have extensive third party reach through the deployment of advertising, analytics, or social sharing functionality. In our four test runs of 100 sites, we saw frequent communications to services that also offered users login accounts for email, social networking, shopping, or news services. In fact, 26 of the 100 most requested domains are operated by such companies (see Table 3).

As noted above, the fact that a site offers login capability does not necessarily mean it correlates or could correlate all data collected across all the domains that it operates with that identifying login information. However, at least some of the domains listed above tie information collection to login identity, and use the third party data for ad targeting purposes [36, 37]. Given the reach of some of these domains, the companies operating them may be able to supplement the information provided to them as a first-party service with significant cross-site and cross-device behavioral data.

5.3 Shared Login Deterministic Tracking

In our study, we also looked for instances of websites sharing personally identifiable information — or hashed personally identifiable information — with third parties. Of the 100 sites we reviewed, we detected 16 sites sharing an email address or user name — or a common hash of one of those — with a total of 60 third party domains.[16] Because we checked only for non-encrypted values and a limited set of hash approaches, additional sharing may have occurred as well.[17]

Although this information could potentially be used to correlate users across different devices, we could not definitively determine the purpose of the information sharing in most cases, and it would be infeasible to do so given the information available. In some cases, the disclosure may be unintentional: if a website structured its site to include personal information in the site's URLs for logged-in users, that data could be passed along in ad requests as part of the referer header.[18] Alternatively, the data could be passed to a third-party service provider with no independent right to use the data, for purposes wholly unrelated to cross-device tracking.

In at least some cases, it seems likely that the data was exchanged to facilitate ad tracking or targeting, though it is not clear if this encompasses cross-device ad tracking. For example, one site shared MD5 and SHA1 hashes of our test email address with a total of 36 different domains. The site's privacy policy explicitly reserved the right to "pass an encrypted or 'hashed' (non-human readable) identifier corresponding to your email address to a Web advertising partner" in order to "enable more customized ads, content or services to be provided to you." This paper demonstrates a lower bound for this

16 We visually inspected the domains to exclude domains that were operated by the first-party publisher. The 60 figure represents the total number of domains after excluding commonly operated domains.

17 Our searches were limited to unkeyed/unsalted MD5, SHA-1, SHA-256, or base64 encoding of our test username and email addresses.

18 When your browser connects to a new domain, the identity of the domain which linked to the new domain is often included as part of the communication. If that "referring" web address included personal information (e.g., http://www.domain.com/username), that information might be passed to the operator of the new domain.

Table 4. Top 20 third-party domains sharing cookie ID values with other parties

Sharing Domain	Times cookie ID shared with third party
demdex.net	150
brtroll.com	91
adsrvr.org	79
w555c.net	70
bidswitch.net	55
criteo.com	40
tidaltv.com	38
addthis.com	31
crwdcntrl.net	30
sitescout.com	30
vindiciosuite.com	30
adsymptotic.com	28
veruta.com	23
mybuys.com	22
contextweb.com	19
steelhousemedia.com	19
everesttech.net	17
gravity.com	16
wtp101.com	14
zdbb.net	12

Table 5. Top 20 third-party domains receiving other domains' cookie ID values

Shared-with Domain	Number of unique domains sending cookie ID values
rubiconproject.com	41
doubleclick.net	30
rlcdn.com	29
adnxs.com	27
bluekai.com	26
casalemedia.com	23
pubmatic.com	23
exelator.com	20
lijit.com	20
demdex.net	19
addthis.com	18
bidswitch.net	18
contextweb.com	18
advertising.com	16
liverail.com	16
visualdna.com	16
krxd.net	14
openx.net	13
adtechus.com	12
yahoo.com	12

activity, as we were unable to detect data exchange if sites used other means to hash or obscure identifying information.

5.4 Graph Sharing

Finally, companies may not generate cross-device graphs themselves. Instead, they may send cookie values to a cross-device tracking company through "cookie syncing"; the cross-device company could return a list of devices it believes to be linked to the same user (see supra 3.2.3).

As shown in Table 4, we detected numerous instances of third-party domains transmitting their own unique cookie IDs to other third parties. In all, we saw 106 domains' cookie IDs transmitted to 210 different domains. These IDs were just the ones flagged by the cookie syncing script described supra, 4.2; the number of actual instances of cookie syncing may well be higher. Nevertheless, the data reveals expansive use of cookie syncing among third parties. As shown in the charts below, one domain, demdex.net, sent a unique identifier to a third party 150 times during our four test runs. The third party advertising technology domains that most frequently received others' unique IDs were rubiconproject.com, doubleclick.net, and rlcdn.com (see Table 5).

As with the sharing of hashed (or unhashed) personally identifiable information, it is not clear whether these incidences of cookie syncing are being done to enable cross device tracking (here, the sharing of device graphs). Cookie syncing is a common practice in the industry that enables various entities in the ad ecosystem to bid more effectively and efficiently on real time ad auctions. It is possible that the primary purpose of these synchronizations is to enable real-time bidding, though there is no obvious way for an end user to discern the purpose of particular cookie value exchanges. However, we detected 20 domains syncing their cookies with companies that specialize in probabilistic cross-device tracking.

6 Transparency

In addition to our four test runs of 100 websites, we looked at the privacy policies of those 100 sites[19] to eval-

19 In some cases, a domain appeared more than once within the list of top sites. For example, bbc.co.uk/news appeared in the top 20 sites for News, and bbc.co.uk/sport appeared in the top 20 sites for Sports. In our data analysis, we count these sites as two sites, not one. Moreover, several of the sites had common ownership — e.g., both Mapquest (top 20 for Reference) and the Huffington Post (top 20 for News) are owned by Verizon, and link to the same privacy policy. We also counted these as two sites (and privacy policies), not one.

uate what disclosures were made to users about cross device tracking, or about data collection and sharing that could enable cross-device tracking. Our review indicated that there was very little explicit disclosure to consumers about cross-device tracking. In fact, several privacy policies we reviewed had been last updated several years ago, before third party advertising and analytics companies began to seriously explore cross-device tracking. Our primary findings are:

- 96 of 100 sites we reviewed disclose that they collect information such as login credential or email addresses from users.
- Only three sites provided specific information to users about enabling third-party cross device tracking.
- 73 sites reserved broader rights to use and share "non-personally identifiable information" such as IP address and cookie IDs compared to "personally identifiable information" (such as names and email addresses).[20]
- 67 sites provided links to industry self-regulatory controls to limit the use of behavioral data for ad targeting; however, few provided information about how consumers could prevent cross-device tracking.

6.1 The vast majority of the sites that we tested collect personally identifiable information such as login or email address

96 of the 100 sites we reviewed offer users the capability of logging in to a persistent account. In many cases, there are obvious benefits to providing a persistent identifier to the site, such as accessing a fantasy team or favorite team scores on sports sites, or saving credit card and shipping information on shopping sites. Many of these sites make logging in easier by integrating with social login services such as those offered by companies including Facebook and Google.

Sites that collect login information may facilitate deterministic cross-device tracking, either by offering third-party functionality on other websites, or by sharing login or hashed login credentials with other third-party data brokers. Sharing of data based on persistent identifiers such as an email address may not always take place in a user's browser; instead, companies could match cross-device data offline in ways that are not observable to consumers. However, few privacy policies provided clear information about whether this was permitted or envisioned.

6.2 Very few sites explicitly discuss third-party cross-device tracking

Only three of the 100 sites we tested linked to a privacy policy that explicitly discussed enabling third parties to engage in cross-device tracking. One of the 100 tested sites stated that:

> Ad Partners may place cookies, web beacons and/or other data collection technologies on the Services to (among other purposes) track how the Services are used, where users go and what they do after they leave the Services; **link users' devices**; and serve more relevant ads on the Services or other websites that you visit. [emphasis added]

Two other commonly owned sites linked to a privacy policy that stated that the company used cookies and similar technologies such as a "Device Graph" which was subsequently defined as "techniques using IP addresses, mobile technologies, and proprietary methods to determine if one or more devices may relate to the same user."

The other privacy policies we reviewed did not make explicit reference to the possibility of cross-device tracking. A substantial number discuss the notion of third-party data collection for advertising, analytics, or social sharing, and many raise the possibility that consumers might be tracked from site to site (though not from device to device). Many policies also direct users to resource pages for third-party advertising self-regulatory groups such as the Network Advertising Initiative (NAI) and the Digital Advertising Alliance (DAA) — or to specific third party providers such as Google — for more information.

6.3 Distinction between personally identifiable information and non-personally identifiable information

73 of the 100 website privacy policies we looked at reserved considerably broader rights to use and share "non personally identifiable information" — such as cookies

20 The FTC regards data as "personally identifiable," and thus warranting privacy protection, when it can be reasonably linked to a particular person, computer, or device. In many cases, persistent identifiers — such as device identifiers, MAC addresses, static IP addresses, or cookies — meet this test [2, 38].

and IP addresses — compared to "personally identifiable information" such as name and email addresses.[21] This distinction likely reflects traditional models of third party online advertising that are based upon cross-site collection correlated to device-specific cookies. However, as noted supra 3.1, this same data could be used for probabilistic cross-device correlation as well, by — for example — looking for devices that share IP addresses during certain periods of the day. The notion of correlating devices based on shared IP addresses was not discussed in the privacy policies we reviewed.

Even for those sites that put in place more rigorous prohibitions on sharing personally identifiable information with third parties, it was not clear if this language would bar all deterministic data sharing methods. For example, it is not clear whether hashed identifiers are considered personally identifiable under many policies' definitions [30]. Twelve of the 100 privacy policies made reference to sharing hashed identifiers, though it is not clear that the purpose is to enable cross-device tracking. (In at least some cases, the language implies that the hashed identifiers are being exchanged in order to match the user with demographic attributes held by a data broker.) Six policies explicitly stated that the company was sharing hashed identifiers with Facebook to enable the company to display ads on Facebook's platform.[22]

Also, most sites that broadly prohibited sharing personally identifiable information with advertisers typically reserved rights to share this information with "service providers" operating on their behalf, or, less frequently, with more vaguely defined entities such as "business partners." The degree of constraints on what such providers or partners could do with the data varied, but most of the language we reviewed could arguably be interpreted to permit sharing identifiers such as email address with a third party in order to display ads on behalf of the site on other browsers or devices.

6.4 Controls

The website privacy policies we reviewed provided little — if any — information about consumer controls to prevent or limit cross-device tracking. 67 out of the 100 policies referred to controls offered by self-regulatory organizations such as NAI, TRUSTe, or the DAA. Those controls allow users to set opt-out cookies that marketers can recognize as an instruction to not use behavioral data for ad targeting. The DAA released formal guidance specifically on cross-device tracking in November 2015. When those rules go into effect in 2017, [41] third-party cross-device tracking companies will be required to disclose in their privacy policies that data may be used across devices. Also, consumers who take advantage of an industry opt-out program on one device should not have that data be used for ad targeting on another device, or have data from another device influence ad targeting on a device where the consumer opts out [39]. However, industry controls do not necessarily limit cross-site or cross-device tracking for purposes beyond ad targeting, such as ad reporting, research, testing, analytics, or security.

Moreover, industry self-regulatory controls only apply to members of self-regulatory organizations; as noted above, in our test runs, our computers connected to 1130 distinct domains, many of which were operated by companies not participating in self-regulatory organizations or global opt-out programs. As of August 2016, the DAA represented that they offer opt-outs for 126 companies; the NAI 99. However, several of the most frequent third-party domains were not covered by one or both programs. Of the top ten third-party services detected in our study, the DAA opt-out only applied to six; the NAI opt-out five.

One half of the sites we reviewed referenced the "Do Not Track" control that is available to signal a preference to limit third party data collection in most browsers. Of these, twenty-two of the sites simply stated that they do not honor the setting and twenty-six stated that the Do Not Track standard was still being worked out by industry.[23] Fifty make no mention whatsoever of Do Not Track. Only two stated that they change site

21 The precise terminology and formulation varied from policy to policy, and in many cases the language could be subject to varying interpretations. This assessment represents our best-faith effort at ascribing the intended meaning of these policies.

22 Many sites' policies noted the presence of social sharing widgets on their sites. Some explicitly stated that those social sites would be able to record your browsing activity if you were logged in; others simply mentioned that the their data collection practices were subject to their privacy policies.

23 The World Wide Web Consortium submitted its two Do Not Track standards for Candidate Recommendation on August 20, 2015 and April 16, 2016, respectively, indicating that the standards are ready for implementation (https://www.w3.org/2011/tracking-protection/). According to the W3C, it publishes a Candidate Recommendation to indicate that the document is believed to be stable and to encourage implementation by the de-

behavior in response to Do Not Track requests — Wiktionary (which treats everyone as if they have selected Do Not Track) and Stumbleupon. Since our review was completed, at least one other site in our test group — reddit.com — has revised its privacy policy to state that it too will limit data sharing in response to Do Not Track signals.[24]

Consumers who wish to prevent or more restrictively limit cross-device tracking may resort to other means, such as the use of tracker blocking software [23, 40]. None of the sites' policies we reviewed made reference to these tools. Tracker blocker programs block connections to domains that are believed to be engaged in cross-site tracking, though the different tools use varying methods in making that determination. Some of these services will whitelist tracking companies that agree to limit data retention periods.[25] Consumers can also configure their browsers to delete or block cookies, though this may not prevent companies from using other identifiers, such as IP address or local storage, to keep state on the user over time [20]. For consumers who block tracking companies or cookies, some websites may refuse to provide content, as their sites are funded by third-party advertising or rely on third-party domain cookies to function properly. Blocking may also impact non-advertising functionality if the site uses different domains to provide different features on the site. Consumers will have to make a choice whether to change their settings, manually whitelist certain websites, or open the website's content in another browser that allows tracking.

Even blocking third party connections may not necessarily prevent all third-party cross-device linkage. Whenever a consumer connects to a website, that website will have access to her IP address, basic information about her browser and computer, and whatever personally identifying information she directly provides. It is theoretically possible that the website could try to match this data with data collected by third parties on other devices. Restricting third-party data collection through the browser however does make this considerably more difficult in practice. Use of a virtual private network (VPN) or the Tor browser would offer additional protection against linkability, though at a cost to performance (and in the case of a VPN, the cost of the service itself).

7 Limitations

While our study demonstrates data collection that could be used to correlate user activity across devices, it is difficult to state with certainty when companies use this data for cross-device tracking purposes. Companies make the determination that two devices are linked on their own servers which cannot be observed by consumers. In the case of probabilistic cross-device tracking, the necessary data collection may not be materially different from the data typically collected by third-party advertising companies — namely IP address, a unique cookie, user agent information, and cross-site behavioral data.[26] It may be impossible to tell whether a company that collects data about a user across different devices has made a determination that two devices are related based on shared IP address, browsing behavior, or any other factor.

Deterministic tracking requires the collection of a persistent cross-device identifier (such as a username or email address), but even when users provide that information to a service, it is not always clear how that information is being logged, used, or shared. Login sites that provide functionality on other sites may architect their systems not to log third-party information in a way that can be easily correlated with login data (and/or they may have strict policy measures in place to restrict such correlation).

For shared login credential cross-device tracking, our study may be both over-inclusive as well as under-inclusive. In a number of cases, we determined that first-party sites were sharing login credentials (or hashed login credentials) with third party companies. However, it is not necessarily clear in these cases that the sharing was done for the purpose of cross-device tracking. Instead, the information could be shared with a data broker for the purpose of looking up demographic data

veloper community (https://www.w3.org/Consortium/Process/Process-19991111/tr.html).

24 Notwithstanding first-party representations, some third parties (such as Twitter, https://support.twitter.com/articles/20169453) independently state that they modify data collection and use practices in response to "Do Not Track" signals.

25 For example, programs like Disconnect.me (https://disconnect.me/help) and Privacy Badger (https://www.eff.org/privacybadger) allow connections to third party services that represent that they honor users' Do Not Track settings.

26 Probabilistic cross-device tracking companies also leverage location information, but that information is less often requested by web publishers than in the mobile app environment.

about the user in order to target ads. Discerning cross-device tracking activity is complicated by the fact that some of the companies who collect login data for cross-device correlation for some clients also provide demographic enrichment data services.[27]

While we were able to detect the sharing of identifiers or common hashes of those identifiers in several cases, we were not able to detect other methods of exchanging identifying information based on login credentials. Companies could use different hashing or obscuring methods, or could hash *salted* data in ways that were undetectable to us.[28] In our study, we detected numerous instances of companies sending unique identifiers (cookie values) to other companies; it is possible that these identifiers were reasonably linked or linkable to personally identifiable information in ways that we did not detect. Thus, the number of instances of sharing login data could be substantially higher than our study indicates. Alternatively, cross-device matching could happen outside of the browser — companies could simply directly exchange information they have about a particular device or individual.[29]

Our study only looked at the possibility of linking two browsers on two virtual computers. We did not study the ability of companies to track users across mobile applications or other devices leveraging static identifiers such as IDFAs or Android IDs. Cross-device tracking in that context could well be easier than in the cross-browser context, as identifiers are persistent across applications (eliminating the need for cookie syncing) and third parties have greater access to information such as geolocation which could improve probabilistic matching. Nor did we look at tracking across connected IoT devices such as smart TVs, gaming consoles, or appliances.[30] We also did not look at network-level tracking by an internet service provider that may provide internet connectivity to multiple devices.

Memory resource issues with our VMs may have affected the data. Our VMs had initial memory resource issues, which may affect data based on latency or other network-based indicators that web developers may monitor and alter activities when serving a particular site. Our data was collected over the course of two months. While we took measures to ensure the browser profiles were fresh by checking logged in sites before importing the profile from the first test phase to the second, the timeframe between the first set of runs (Run 1 and Run 2) and the second (Run 3 and Run 4) may have affected that data. Additionally, having only four total runs, we may not have sufficiently broad coverage to identify the full scope of cross-device tracking. In particular, probabilistic cross-device tracking practices may require a more constant and continual pattern of activity.

Finally, because of human error in coding our test runs, we did not collect data on two sites: EA.com and Gamespot.com. As a result, we missed further opportunities to detect additional data collection and sharing, and artificially lowered our average and median results marginally.

8 Conclusion

Our research demonstrates that websites share extensive data with third party services that could allow those third parties to track user behavior across multiple devices, and consumers lack the necessary information to determine precisely whether and when this information is used for cross-device tracking. However, we were unable to conclusively determine how often the data was shared for that purpose. Data that could be used for

27 For example, a first-party publisher can look up an email address with a data broker service in order to get demographic information associated with that individual. In that case, the first party would be sharing the data — or a hash of that data — with a third party in order to enrich its data profile of the individual rather than to facilitate cross-device tracking.

28 A "cryptographic salt" is data (often a random string) appended to a sensitive data value (e.g., password) before hashing, with the "salted hash" and the salt value being stored in a database. While it is extremely challenging to reverse a robust hash function, an attacker could attempt to guess what data yielded a given hash by creating a list of possible values and their respective hashes, looking for a match. By adding the salt, such an attacker would need to greatly increase the number of possible values to guess in order to find the correct one, potentially rendering such an attack infeasible. Meanwhile, the company who has the original data knows the salt, and can recalculate the hash of the salted data easily.

29 Companies exchanging data offline would need access to a shared identifier for each individual or device. That could include real-world identifying information, a persistent device-specific identifier such as an IDFA, or cookie IDs if the companies had previously synced cookies.

30 Further, we did not look at cross-platform linking on the same device — e.g., tracking across browsers on one mobile device, or correlating app usage and browsing data on the same device. Although this is not technically "cross-device" tracking, it does raise many of the same issues and challenges. However, at least several companies have publicly purported to be engaged in cross-device, third-party tracking [34–37].

probabilistic tracking — IP address, cookies, location, and behavioral data — is the same data that is routinely collected and logged for any third-party advertising, even non-targeted advertising. Login sites with a broad third-party presence could configure their systems to not collect and store data in a way that could be easily correlated with login identity, or they could have policy restrictions against correlation. Sites that shared identifiers or hashed identifiers with third parties could be sharing data for other purposes, such as onboarding demographic information or uploading data to a data management service provider. It is possible that limited third-party cross-device tracking is happening today, though any retained data could be used for ex post cross-device correlation in the future unless there are contractual prohibitions on this usage. Future research may attempt to characterize or quantify when content on one device is personalized based on conduct on another device (such as ads that are retargeted across devices).

Our review of 100 publisher privacy policies did not provide substantial clarity about the extent of third-party cross-device tracking. Many websites reserved broad rights to share non-personally identifiable information with advertising and analytics companies, though cross-device tracking was very rarely identified as a potential purpose. Many sites also reserved more limited rights to share personally identifiable information with partners, but it was often difficult to interpret what cross-device tracking using such information was permitted under the terms of their disclosures. We did not review the privacy disclosures of the embedded third-party companies; that may be a useful avenue for future research.

References

[1] Federal Trade Comm'n, "Privacy Online: Fair Information Practices in the Electronic Marketplace," https://www.ftc.gov/sites/default/files/documents/reports/privacy-online-fair-information-practices-electronic-marketplace-federal-trade-commission-report/privacy2000.pdf, May 2000.

[2] Federal Trade Comm'n, "Self-Regulatory Principles for Online Behavioral Advertising," https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-staff-report-self-regulatory-principles-online-behavioral-advertising/p085400behavadreport.pdf, Feb. 2009.

[3] Federal Trade Comm'n, "Protecting Consumer Privacy in an Era of Rapid change: Recommendations for Businesses and Policymakers (Privacy Report)," https://www.ftc.gov/sites/default/files/documents/reports/federal-trade-commission-report-protecting-consumer-privacy-era-rapid-change-recommendations/120326privacyreport.pdf, March 2012.

[4] Fed. Trade Comm'n, "FTC Settlement Puts an End to 'History Sniffing' by Online Advertising Network Charged With Deceptively Gathering Data on Consumers," https://www.ftc.gov/news-events/press-releases/2012/12/ftc-settlement-puts-end-history-sniffing-online-advertising, Dec. 2012.

[5] Fed. Trade Comm'n, "Online Advertiser Settles FTC Charges ScanScout Deceptively Used Flash Cookies to Track Consumers Online," https://www.ftc.gov/news-events/press-releases/2011/11/online-advertiser-settles-ftc-charges-scanscout-deceptively-used, Nov. 2011.

[6] Fed. Trade Comm'n, "FTC Puts an End to Tactics of Online Advertising Company That Deceived Consumers Who Wanted to 'Opt Out' from Targeted Ads," https://www.ftc.gov/news-events/press-releases/2011/03/ftc-puts-end-tactics-online-advertising-company-deceived, March 2011.

[7] TRUSTe/Nationnal Cyber Security Alliance, "U.S. Consumer Privacy Index 2016," https://www.truste.com/resources/privacy-research/ncsa-consumer-privacy-index-us/.

[8] M. Madden, L. Rainie, "Americans Attitudes About Privacy, Security, and Surveillance," http://www.pewinternet.org/2015/05/20/americans-attitudes-about-privacy-security-and-surveillance/, May 2015.

[9] I. Altaweel, N. Good, C.J. Hoofnagle, "Web Privacy Census," Technology Science, http://techscience.org/a/2015121502/, Dec. 2015.

[10] S. Englehardt, D. Reisman, C. Eubank, P. Zimmerman, J. Mayer, A. Narayanan, E. Felten, "Cookies That Give You Away: The Surveillance Implications of Web Tracking," in *Proceedings of the 24th International conference on World Wide Web,* (New York, NY, USA), pp. 289–299, ACM, 2015.

[11] D. Malandrino, A. Petta, V. Scarano, L. Serra, R. Spinelli, B. Krishnamurthy, "Privacy Awareness about Information Leakage: Who Knows What about Me?," in *Proceedings of the 12th ACM Workshop on privacy in the electronic society,* (New York, NY, USA), ACM, pp. 279–284, 2013.

[12] J. Mayer, "Tracking the Trackers: Where Everybody Knows Your Username," http://cyberlaw.stanford.edu/blog/2011/10/tracking-trackers-where-everybody-knows-your-username, Oct. 2011.

[13] J. Zang, K. Dummit, J. Graves, P. Lisker, L. Sweeney, "Who Knows What About Me? A Survey of Behind the Scenes Personal Data Sharing to Third Parties by Mobile Apps," Technology Science, http://jots.pub/a/2015103001/, Oct. 2015.

[14] J. Ren, A. Rao, M. Lindorfer, A. Legout, D. Choffnes, "ReCon: Revealing and Controlling PII Leaks in Mobile Network Traffic," in *International Conference on Mobile Systems, Applications, and Services,* (Singapore), ACM, June 2016.

[15] S. Son, D. Kim, V. Shmatikov, "What Mobile Ads Know About Mobile Users", in *Network and Distributed System Security Symposium,* (San Diego, CA, USA), Internet Society, Feb. 2016.

[16] A. Razaghpanah, N. Vallina-Rodriguez, S. Sundaresan, C. Kreibich, P. Gill, M. Allman, V. Paxson, "Haystack: A Multi-

Purpose Mobile Vantage Point in User Space," https://www.ftc.gov/system/files/documents/public_comments/2016/10/00038-129144.pdf, Oct. 2016.
[17] A. McDonald, L. Cranor, "The Cost of Reading Privacy Policies," in *Journal of Law and Policy for the Information Society*, 2008.
[18] J. Graves, "An Exploratory Study of Mobile Application Privacy Policies," Technology Science, http://techscience.org/a/2015103002/, Oct. 2015.
[19] A. Soltani, S. Canty, Q. Mayo, L. Thomas, C.J. Hoofnagle, "Flash Cookies and Privacy," https://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862, Aug. 2009.
[20] G. Acar, C. Eubank, S. Englehardt, M. Juarez, A. Narayanan, C. Diaz, "The Web Never Forgets: Persistent Tracking Mechanisms in the Wild," in *Proceedings of the 2014 Conference on Computer and Communications Security*, (Scottsdale, AZ, USA), ACM, pp. 674–689, 2014.
[21] N. Kapravelos, A. Kapravelos, W. Joosen, C. Kruegel, F. Piessens, G. Vigna, "Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting," https://www.cs.ucsb.edu/~vigna/publications/2013_SP_cookieless.pdf, 2013.
[22] R. Díaz-Moralesl, "Cross-Device Tracking: Matching Devices and Cookies," https://arxiv.org/pdf/1510.01175.pdf, Oct. 2015.
[23] Fed. Trade Comm'n, "Consumer Information: Online Tracking," https://www.consumer.ftc.gov/articles/0042-online-tracking.
[24] P. Eckersley, "How Unique is Your Web Browser?," https://panopticlick.eff.org/static/browser-uniqueness.pdf.
[25] J. Angwin, "Meet the Online Tracking Device That is Virtually Impossible to Block," https://www.propublica.org/article/meet-the-online-tracking-device-that-is-virtually-impossible-to-block, July 2014.
[26] L. Olejnik, C. Castelluccia, A. Janc, "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.
[27] R. Bilton, "Cross-device tracking, explained," http://digiday.com/publishers/deterministic-vs-probabilistic-cross-device-tracking-explained-normals/, Aug. 2015.
[28] L. Lessig, Code Version 2.0., pp. 33–35, 2006.
[29] M. Cavna, "NOBODY KNOWS YOU'RE A DOG': As iconic Internet cartoon turns 20, creator Peter Steiner knows the idea is as relevant as ever, Washington Post," https://www.washingtonpost.com/blogs/comic-riffs/post/nobody-knows-youre-a-dog-as-iconic-internet-cartoon-turns-20-creator-peter-steiner-knows-the-joke-rings-as-relevant-as-ever/2013/07/31/73372600-f98d-11e2-8e84-c56731a202fb_blog.html, July 2013.
[30] E. Felten, "Does Hashing Data Make Data 'Anonymous'?," https://www.ftc.gov/news-events/blogs/techftc/2012/04/does-hashing-make-data-anonymous, Apr. 2012.
[31] Fed. Trade Comm'n, "FTC Isssues Warning Letters to App Developers Using 'Silverpush' Code," https://www.ftc.gov/news-events/press-releases/2016/03/ftc-issues-warning-letters-app-developers-using-silverpush-code, March 2016.
[32] S. Englehardt and A. Narayanan, "Online tracking: A 1-million-site measurement and analysis," http://randomwalker.info/publications/OpenWPM_1_million_site_tracking_measurement.pdf, May 2016.
[33] , "Infographic: Everything You Need to Know about Real Time Bidding for Display Ads," http://marketingland.com/infographic-real-time-bidding-83186, May 2014.
[34] Tapad, "Who We Are," http://www.tapad.com/about-us/who-we-are/.
[35] Drawbridge, "The Largest Independent Cross-Device Identity Solution," https://drawbridge.com/c/graph.
[36] Facebook, "What information does Facebook get when I visit a site with a Like button?," https://www.facebook.com/help/186325668085084.
[37] B. Barrett, "Uh Oh: Google Expands its Ad Tracking. But, Yay: It's Opt-in," https://www.wired.com/2016/06/latest-ad-tracking-move-google-gets-opt-right/, June 2016.
[38] J. Rich, "Keeping Up with the Online Advertising Industry," https://www.ftc.gov/news-events/blogs/business-blog/2016/04/keeping-online-advertising-industry, 21 April 2016.
[39] Digital Advertising Alliance, "Application of the DAA Principles of Transparency and Control to Data Used Across Devices," http://www.aboutads.info/sites/default/files/DAA_Cross-Device_Guidance-Final.pdf, Nov. 2015.
[40] L. Schifferle, "Online tracking — more than cookies," https://www.consumer.ftc.gov/blog/online-tracking-more-cookies, June 2016.
[41] L. Tonsager, "Digital Advertising Alliance Will Begin Enforcing its Cross-Device Guidance February 1, 2017," https://www.insideprivacy.com/advertising-marketing/digital-advertising-alliance-will-begin-enforcing-its-cross-device-guidance-february-1-2017/, Oct. 2016.