Altaweel I, Good N, Hoofnagle C. Web Privacy Census. Technology Science. 2015121502. December 15, 2015.
http://techscience.org/a/2015121502


# Web Privacy Census

Ibrahim Altaweel, Nathaniel Good, and Chris Jay Hoofnagle

## Highlights

- We repeated a 2012 survey of tracking mechanisms such as HTTP cookies, Flash cookies, and HTML5 storage, used by top 25,000 most popular websites
- We found that the top 100 most popular sites would collect over 6,000 HTTP cookies with 83% being third-party cookies
- We found that Google tracking infrastructure is on 92 of the top 100 most popular websites and on 923 of the top 1,000 websites, providing Google with a significant surveillance infrastructure online

| Description | Date | Type | Sites Crawled | Total HTTP Cookies | First-Party HTTP Cookies | Third-Party HTTP Cookies | Sites Using Flash Cookies | Sites Using HTML5 Local Storage |
|---|---|---|---|---|---|---|---|---|
| **Top 100 Sites** | 2015-07-01 | shallow | 100 | 6,280 | 1,091 (17%) | 5,189 (83%) | 5 (5%) | 63 (63%) |
| **Top 100 Sites** | 2015-07-01 | deep | 100 | 1,2857 | 1,265 (10%) | 11,592 (90%) | 10 (10%) | 76 (76%) |
| **Top 1,000 Sites** | 2015-07-01 | shallow | 1,000 | 80,821 | 10,374 (13%) | 70,447 (87%) | 39 (4%) | 613 (61%) |
| **Top 1,000 Sites** | 2015-07-01 | deep | 1,000 | 134,769 | 10,871 (8%) | 123,898 (92%) | 62 (6%) | 649 (65%) |
| **Top 25,000 Sites** | 2015-07-01 | shallow | 25,000 | 1,065,076 | 135,767 (13%) | 929,309 (87%) | 585 (2%) | 8,688 (35%) |

*Overall summary of results for shallow and deep crawls for the top 100, 1,000 and 25,000 websites*

## Abstract

Most people may believe that online activities are tracked more pervasively now than they were in the past. In 2011, we started surveying the online mechanisms used to track people online (e.g., HTTP cookies, Flash cookies and HTML5 storage). We called this our Web Privacy Census. We repeated the study in 2012. In this paper, we update the study to 2015.

**Results summary:** Our approach uses web crawler software to simulate online browsing behavior, and we record the occurrences of tracking mechanisms for the top 100, 1,000, and 25,000 most popular websites. We found that users who merely visit the homepages of the top 100 most popular sites would collect over 6,000 HTTP cookies in the process (see Top 100 Websites - Shallow Crawl). Eighty-three percent of cookies are third-party cookies. The homepages of popular sites placed cookies for 275 third-party hosts. In just visiting the homepage of popular sites, we found 32 websites placed 100 or more cookies, 7 websites placed 200 or more cookies, and 6 websites placed 300 or more cookies. We found that Google tracking infrastructure is on 92 of the top 100 most popular websites and on 923 of the top 1,000 websites. This means that Google's ability to track users on popular websites is unparalleled, and it approaches the level of surveillance that only an Internet Service Provider can achieve.

## Introduction

How is online privacy doing? Public policy makers regularly propose measures to give consumers more privacy rights online. These measures rely on the assumption that the web privacy landscape has become worse for consumers and that online tracking is more pervasive now than in the past. As policymakers consider different approaches for addressing Internet privacy, it is critical to understand how interventions such as negative press attention, self-regulation, Federal Trade Commission enforcement actions, and direct regulation affect tracking.

In 2011, we began taking comprehensive measures of online privacy. We term our measures the Web Privacy Census. We took a Web Privacy Census in 2011 [1] and 2012 [2]. In this paper, we report on the Web Privacy Census of 2015.

The earlier Web Privacy Censuses measured how much information could be associated with a visitor to a website. Tracking activities relied on cookies, Flash cookies and HTML5 storage.

A cookie is a message a web browser (e.g., Internet Explorer, Safari, or Firefox) stores when a website it visits requests it to do so. The browser sends the message back to the server each time the browser requests a page from the server. Websites often use cookies to track visits to the same or different websites. A third-party cookie is one that appears in your browser when you visit a web page even though the cookie is not specific to the website you visited.

Flash cookies, more formally termed Local Shared Objects (LSOs), are like regular cookies except that they do not appear in a browser’s list of cookies, making them harder to detect and delete.

HTML5 is a markup language used for presenting content on web pages. Web browsers that support HTML5 also allocate some local storage in the browser to store data. Browsers store cookies in the local storage, for example. However, storage of larger amounts of data is

allowed. The size of a cookie should not exceed 4093 bytes or 4K, while a Flash cookie is 100KB. The HTML5 storage limit is far larger (at least 5MB).

Our Web Privacy Census found a marked increase in HTML5 storage usage and a sharp decline in Flash cookies between 2011 and 2012. An increase in HTML5 storage does not directly correlate with an increase in tracking, as an HTML5 storage object can hold any information that the developer needs to store locally. However, this information can potentially contain information used to track users and can persist. Our 2015 census found that regular cookie counts continue to increase, with larger and larger numbers of third-party cookies in use Cookies are present on every website in the top 100 most popular websites, with approximately 34% using HTML5 storage, more than double the amount we counted in 2011.

## Background

As early as 1995, Beth Givens of the Privacy Rights Clearinghouse suggested that federal agencies create benchmarks for online privacy. The first attempts at web measurement found relatively little tracking online in 1997: only 23 of the most popular websites used cookies on their homepages [3]. But within a few years, tracking for network advertising appeared on many websites. By 2011, all of the most popular websites employed cookies. Below is a historical summary. Table 1 presents a reverse timeline.

- The Electronic Privacy Information Center made the earliest attempts to enumerate privacy practices in a systematic fashion. In June 1997, it released “Surfer Beware: Personal Privacy and the Internet,” a survey of the top 100 websites. Only 17 of the top 100 websites had privacy policies. Twenty-three sites used cookies. This observation may underrepresent the actual number of sites using cookies. It appears that EPIC used a “surface crawl” to detect those cookies, meaning that it only visited the homepage of the site and did not click other links. By 2009, Soltani et al. found cookies on 98 of the top 100 sites, and by 2011, Ayenson et al. found cookies on all 100 most popular sites [1] (see discussion below).

- In “Surfer Beware II: Notice is Not Enough, published in June 1998”, EPIC surveyed websites of companies that had recently joined the Direct Marketing Association [4]. At the time, the Direct Marketing Association (DMA) committed to basic privacy protections, including notice and an ability for consumers to opt out. EPIC found 76 new members of the DMA, but only 40 had websites. Of those 40, all collected personal information. Only eight of the sites had a privacy policy.

- The Federal Trade Commission conducted the first large-scale privacy measurement study in “Privacy Online: A Report to Congress,” released in June 1998. The Commission examined the privacy practices of 1,402 websites using a sophisticated sample procedure to ensure that a variety of consumer-oriented websites were

studied (health, retail, financial, sites directed at children, and the most popular websites). The FTC found that

"The vast majority of Web sites—upward of 85%—collect personal information from consumers. Few of the sites—only 14% in the Commission's random sample of commercial Web sites—provided any notice with respect to their information practices, and fewer still—approximately 2%—provided notice by means of a comprehensive privacy policy." [5]

- In EPIC's "Surfer Beware III: Privacy Policies without Privacy Protection," the group surveyed the practices of 100 ecommerce sites [6]. This 1999 report was the most comprehensive but also the last of the EPIC surveys. It evaluated sites for compliance with a full range of fair information practices, such as whether the site collected personal information, whether the site linked to a privacy policy, whether the site agreed to a seal program, and whether users had access and correction rights for personal information. Eighty-six of the sites used cookies, 18 lacked privacy policies, and 35 had some form of network advertiser active on the site. The text of the report makes it clear that EPIC evaluated both the privacy politics of these sites and tested them to see whether they set cookies. However, it is unclear whether EPIC performed a surface crawl of just the homepage or a deeper crawl that explored more of the site.

- In May 2000, the Federal Trade Commission released a survey of sites that detected third-party cookies [7]. In its study, the FTC drew from two groups of websites: those with more than 39,000 visits a month and a second sample of popular sites (91 of the top 100). The FTC found that, "57% of the sites in the Random Sample and 78% of the sites in the Most Popular Group allow the placement of cookies by third parties…. The majority of the third-party cookies in the Random Sample and in the Most Popular Group are from network advertising companies that engage in online profiling."

- In a multiple-year study of 1,200 websites, Bala Krishnamurthy and Craig Wills found increasing collection of information about users from an increasingly concentrated group of tracking companies [8]. Krishnamurthy and Wills describe what we call "DNS aliasing" in their paper (also described in their 2006 paper), a practice where, "…what appeared to be a server in one organization (e.g. w88.go.com) was actually a DNS CNAME alias to a server (go.com.112.2o7.net) in another organization (Omniture)." They found a massive increase in such aliasing: "…the percentage of first-party servers with multiple top third-party domains has risen from 24% in Oct'05 to 52% in Sep'08…This increase is significant because it shows that now for a majority of these first-party servers, users are being tracked by two and more third-party entities." It is also significant because through DNS aliasing, tracking companies can present cookies to users directly as first parties, thereby circumventing third-party cookie blocking. By decoding aliased domains, Krishnamurthy and Wills found that third-party trackers became more concentrated. Sampling from five periods, they found the

concentration grew from 40% in October 2005 to 70% in September 2008. Further, they found that "The overall share of the top-five families: Google, Omniture, Microsoft, Yahoo and AOL extends to more than 75% of our core test set with Google alone having a penetration of nearly 60%."

- In June 2009, Gomez et al. published the KnowPrivacy report. The report focused on several areas of consumer privacy and featured a large-scale crawl of sites based on data from Ghostery [9]. Google-owned trackers were present on over 88% of a sample of 393,829 distinct domains. Further, in a survey of the top 100 sites, Google Analytics appeared on 81 of them.

- In August 2009, Soltani et al. demonstrated that popular websites used "Flash cookies" to track users [10]. Some advertisers adopted this technology because it allowed persistent tracking even where users took steps to avoid web profiling. Soltani et al. also demonstrated "respawning" on top sites with Flash technology. This allowed sites to reinstate HTTP cookies deleted by a user, making tracking more resistant to users' privacy-seeking behaviors. In a survey of the top 100 sites according to Quantcast, Soltani et al. found 3602 cookies set on 98 of the top 100 sites. They also found 281 Flash cookies set on 54 of the top 100 sites.

- In July 2010, Julia Angwin, Tom McGinty, and Ashkan Soltani of the *Wall Street Journal* reported that in a scan of the top 50 sites, 3,180 "tracking files" (comprising HTTP cookies, Flash cookies, and web beacons) were detected [11]. Twelve sites set over 100 each.

- In 2010, Michael Coates surveyed the top 1,000 websites in order to determine how many used HTTPS [12]. Coates sent a basic HTTPS request to these sites, and they responded with 559 cookies. Coates's method appeared to not collect any third-party cookies.

- Flash cookies are now a major focus of research. In 2001, McDonald and Cranor of Carnegie Mellon investigated the presence of Flash cookies on websites [13]. They found a dramatic decline from the Soltani et al. investigation in 2009. McDonald and Cranor found Flash cookies on only 20 of the top 100 sites. They were also careful to attempt to determine whether Flash cookie values were unique or not. Six of the top 100 sites had Flash cookies that were not unique, and thus probably not used to track individuals.

- Krishnamurthy et al. made significant contributions to the study of privacy "leakage." In a study of websites that required registration, they found that a majority of the popular sites they analyzed "directly leak sensitive and identifiable information to third-party aggregators" [14]. The problem they identified was widespread: "56% of the 120 popular sites in our study (75% if we include userids) directly leak sensitive and identifiable in formation to third-party aggregators."

- In July 2011, Stanford Law/Computer Science graduate student Jonathan Mayer released "FourthParty," an "open-source platform for measuring dynamic web content" [15]. Mayer posted the raw data from web crawls made with the platform and released two reports flowing from the system. In the first, Mayer tested how members of the Network Advertising Initiative (NAI) interpret opt-outs [16]. The NAI considers the scope of opt out rights to pertain only to targeting ads, not to tracking. Thus, if a consumer opts out, NAI members may still track them. Mayer found that half of the NAI members tested (N=64) still used tracking cookies despite an opt-out.

- In the second report, Mayer found that in developing FourthParty, he detected "browser history stealing" [17]. This is a practice where a website "exploits link styling to learn a user's web browsing history. The approach is simple: to test whether the user has visited a link, add it to a page and check how it's styled."

- In August 2011, Ayenson et al. surveyed the top 100 websites, simulating a user session by clicking on 10 random links on each site [1]. They detected cookies on all top 100 sites. They found 5,675 cookies, 4,615 of which were set by third parties. They detected 600 third-party hosts. Of the top 100 sites, 97, including popular government website,s used Google-controlled cookies. Ayenson et al. found that 17 sites used HTML5 local storage, and seven of those sites had HTML5 local storage and HTTP cookies with matching values [1]. Flash cookies were present on 37 of the top 100 sites.

- In October 2011, Jonathan Mayer tested signup and interaction on 185 of the Quantcast top 250 sites. He found 113 of the sample leaked user ids or usernames to third parties [18].

- In "Pixel Perfect: Fingerprinting Canvas in HTML5," a study done in 2012 by Mowery and Shacham, the relationship between the web browser and the operating system was investigated in order to understand how each system creates its own fingerprint [19]. Binding the browser with an operating system functionality and hardware allows website to have more information about users. Additionally, Three-dimensional graphics (WebGL) and browser font are used to produce a unique image, which is used as a fingerprint, that can be used to track users online.

- "Understanding What They Do With What They Know," released in 2012 by Wills, et al., investigated what Web advertisers do with information gathered from a user [20]. Advertisements shown to users during experimental controlled browsing sessions and personal interests shown in Ad Preference Managers were analyzed and discussed. The authors found that the Google ad network displays personalized ads, which are categorized in the Ad Preference manager of the user. The ad network uses personal information, including users' private information, in the data collected to generate advertisements in real time. The study also discovered that even though Facebook does not generate ads based on users' browsing behavior on non-Facebook sites, it

uses the Facebook Like button to understand users' interests and show ads based on their interests.

- "FPDetective: Dusting the Web for Fingerprinters," released in 2013 by Acar, discussed how the FPDetective framework detects and analyzes web-based fingerprints [21]. The study also found weaknesses in both the Tor browser and Firegloves, two browsers that pride themselves on concealing fingerprints, that would allow online trackers to identify a user. The authors used FPDetective as a crawler and were able to gather the information to pick up on properties that relate to a user's fingerprint.

- Malandrino, Krishnamurthy et al.'s "Privacy Awareness about Information Leakage: Who Knows About Me?" study considered users' lack of access to and awareness of their private information online [22]. The study compared the amount of sensitive information leaked when using different privacy protection tools, including NoTrace, AdBlock Plus, Ghostery, NoScript, and RequestPolicy. Although they concluded that no privacy extension can fully protect users online, NoTrace was praised for showing users a behind-the-scenes view of the availability of their personal information to trackers.

- Olejnik et al. in "Why Johnny Can't Browse in Peace: On the Uniqueness of Web Browsing History Patterns" investigated how history-based user fingerprinting is done [23]. With a dataset of 300k users' web browsing histories, the pages users visited, and sites they repeatedly returned to, the study found that more than 69% of users have a unique fingerprint. Consequently, web browsing histories can easily be traced to particular users and their personal preferences by web authors.

- Mayer and Mitchell explored third-party tracking and advertising in their study, "Third-Party Web Tracking: Policy and Technology." They used FourthParty, an open-source web platform that measures dynamic web content, to crawl Alexa's Top 500 sites [24]. In the study, Mayer and Mitchell found that of the 11 ad-blocking tools they tested, all blocked third-party advertising. However, the ad-blocking tools did not differentiate between advertising content and advertising-related tracking content. They concluded that without the configuration of options, ad-blocking software can only be slightly effective, and so is primarily a solution for more advanced users.

- In "Privacy and Online Social Networks: Can Colorless Green Ideas Sleep Furiously," Krishnamurthy discussed online social networks (OSNs) and their responsibility, as the parties with the most detail about their users' interactions, to be more transparent about the flow of users' private information to other sites over time [25]. Krishnamurthy believed that with more transparency and tools such as the Facebook extension Privacy IQ, users can get a better understanding of their privacy and what actions they may need to take to attain their preferred level of privacy on social

networks. He suggested that OSNs have the means to bridge the gap between users and privacy protection and should be invested in doing so.

- In “Fast and Reliable Browser Identification with JavaScript Engine Fingerprinting,” Mulazzani et al. also studied how spoofing a user agent string, a string that a browser or other applications generate and send to web servers to identify themselves, does not successfully hide the user’s identity [26]. They tested the underlying JavaScript engine in multiple browsers and browser versions to find that they could reliably determine the user’s browser without regard to the user agent at all.

- In the 2013 study, “Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting,” on web-based device fingerprinting, Nikiforakis et al. of the University of California, Santa Barbara surveyed more than 800,000 users and conducted a 20-page crawl of Alexa’s top 10,000 websites [27]. They found that users who installed browser or user agent-spoofing extensions create a more unique fingerprint for themselves. The study found that the extensions are not able to completely hide the browser’s identity (they are unable to spoof particular methods or properties), resulting in the user being even more recognizable.

- In a 2014 device fingerprinting position paper, “Obfuscation For and Against Device Fingerprinting,” Acar discusses the power and knowledge asymmetry that arises in relation to device fingerprinting because a user has no knowledge of where his or her data is used and no control over how it is gathered [28]. Acar also comments on the uselessness of spoofing user agents as a way to prevent tracking. The conclusion is that more effective tools such as obfuscation with the Tor browser are needed to combat fingerprinting.

- In “Cookies That Give You Away: Evaluating the Surveillance Implications of Web Tracking,” released in 2014, Reisman et al. discovered that multiple web pages with embedded trackers can connect a user’s web page visits back to the specific user [29]. By using simulated browsing profiles, the also discovered that over half of the most popular web pages that have embedded trackers leak a user’s identity to other parties.

- “The Web Never Forgets: Persistent Tracking Mechanisms in the Wild,” a study done in 2014 by Acar et al., focused on a tracking mechanism called canvas fingerprinting [30]. A canvas fingerprint is an image with text that is drawn in the browser and sent to the requesting site the user is on. This type of tracking produces a unique fingerprint without the user being aware, because each system produces a different image. This paper discusses cookie syncing and respawning as tracking techniques to be wary of because they allow domain-to-domain communication and consistent tracking, respectively, after a user wipes their cookies.

| Study | Year | Major Finding | Sample Size |
|---|---|---|---|
| **Acar et al. [30]** | 2014 | A new technique called canvas fingerprinting is used to track users–5.5% of the sample size ran canvas fingerprinting on their homepage. | Top 100,000 sites from the Alexa database |
| **Reisman et al. [29]** | 2014 | Embedded trackers in website allow users to be tracked. | Top 500 Alexa websites |
| **Acar [28]** | 2014 | It discussed findings from [21] indicating that 145 of the top 10,000 websites use Flash-based fingerprinting and 400 of the top one million websites use JavaScript-based fingerprinting. | Top 10,000 websites from the Alexa database |
| **Mulazzani et al. [26]** | 2013 | JavaScript engine fingerprinting is a practical approach to identify and verify an specific browser, even for mobile technologies. | 189 tests |
| **Acar [21]** | 2013 | The FPDetective framework found that 404 sites in the sample size gathered users' fingerprints through their homepages using Javascript-based font probing. | Top million websites from Alexa |
| **Krishnamurthy [25]** | 2013 | Current information protection methods of online social networks (OSNs) are not adequate enough to prevent users' data from being shared by parties across sites. | N/A |
| **Nikiforakis et al. [27]** | 2013 | 40 sites (0.4% of the Alexa top 10,000) are utilizing fingerprinting code from the three commercial providers mentioned in this work. | 20-page crawl of each of the Alexa top 10,000 sites |
| **Malandrino, Krishnamurthy et al. [22]** | 2013 | Aggregators can collect much information about users' online profiles; one of the top ten aggregators in this study collects 87% of a user's private data. | Top 100 sites from 15 Alexa categories |
| **Olejnik et al. [23]** | 2012 | More than 69% of users tested have a unique fingerprint, some larger than 18 bits, just based on their browsing histories. | 368,284 users' web histories |
| **Wills [20]** | 2012 | Advertisements are generated based on details a person's intimate and private life, such as their financial life and sexual orientation. | 15-20 sessions to visit other sites while logged into Facebook |
| **Mowery, Shacham [19]** | 2012 | Revolutionary system to produce fingerprints based on browser font and WebGL rendering. | Samples from 300 distinct members of the Mechanical Turk Marketplace (AI service from |

| Study | Year | Major Finding | Sample Size |
|---|---|---|---|
| | | | Amazon) |
| **Mayer, Mitchell [24]** | 2012 | Ad blocking software is not effective for less advanced users. Out of the 11 ad-blocking tools tested, all blocked third-party advertising but allowed tracking. | 3 consecutive crawls of the Alexa top 500 sites |
| **Mayer [18]** | 2011 | Most popular websites were "leaking" usernames and userids to third parties. | 185 of the Quantcast top 250 |
| **Ayenson et al. [1]** | 2011 | 5,675 HTTP cookies detected, 4,615 of which were third-party. 37 sites with 100 Flash cookies detected. All top websites had cookies. | Top 100 sites, 10-click user session simulated |
| **Mayer [17]** | 2011 | Network Advertising Initiative members continued to use tracking cookies after opt-out. | 64 of the Network Advertising Initiative Members |
| **Krishnamurthy & Wills [8]** | 2011 | Majority of popular websites with registration leaking personal data to third parties. | Array of popular websites that required registration |
| **McDonald & Cranor [13]** | 2011 | Flash cookies present on 20 of top 100 sites. | Surface crawl of homepages of top 100 sites |
| **Coates [12]** | 2010 | 559 first-party cookies detected. | Limited HTTPS request to top 1,000 sites |
| **Angwin et al. (Wall Street Journal What They Know) [11]** | 2010 | 3,180 tracking mechanisms detected. Only one site lacked cookies. | Top 50 sites, 20-click user session simulated |
| **Gomez et al. (KnowPrivacy Report) [9]** | 2009 | Google-owned web beacons present on 88% of a large sample of websites. | 393,829 unique domains |
| **Soltani et al. [10]** | 2009 | 3602 HTTP cookies detected, 281 Flash cookies detected. 98 of the top 100 sites had cookies. | Top 100 sites, 10-click user session simulated |
| **Krishnamurthy et al. [14]** | 2009 | Large increase in DNS aliasing; penetration of major third-party trackers to 75% of sample sites. | 1,200 sites scanned over four years |
| **FTC [5]** | 2000 | 57% of the sites in the Random Sample and | Random sample |

| Study | Year | Major Finding | Sample Size |
|---|---|---|---|
| | | 78% of the sites in the Most Popular Group set cookies. | of 335 sites and 91 of top 100 sites |
| **EPIC Surfer Beware III [6]** | 1999 | 86 used cookies. | 100 e-commerce sites |
| **FTC Privacy Online [7]** | 1998 | Most websites collect personal info, but only 14% have privacy notices. | 1,400 |
| **EPIC Surfer Beware II [4]** | 1998 | Few of the newest DMA members had privacy policies. | New DMA members |
| **EPIC Surfer Beware I [3]** | 1997 | Homepages of 23 sites used cookies. | Top 100 |

**Table 1. Reverse timeline of online privacy measures.**

Since our Web Privacy Census of 2012, online advertising and metrics companies have developed even more sophisticated ways to track and identify individuals online. So, in this study for 2015, we intended to formalize the benchmarking process and measure Internet tracking consistently over time. In this Web Privacy Census, we seek to explore:

- How many entities are tracking users online?
- What vectors (technologies) are most popular for tracking users?
- Is there displacement (i.e., a shift from one tracking technology to another) in tracking practices?
- Is there greater concentration of tracking companies online?
- What entities have the greatest potential for online tracking and why?

## Methods

To answer the questions above, we use a web crawler, a computer program that systematically browses the Internet, to run a crawl on the top 100, 1,000, and 25,000 sites ranked by Quantcast. The crawler determines the number of HTTP cookies, Flash cookies, and HTML5 local storage placed by each website and compares these numbers with results from our 2012 survey. We collect data using deep and shallow crawls within the pages of a domain. Shallow crawls consist of visiting only the homepage of each site, while deep crawls visit the homepage and two other links at random on the site.

We collect data on the top 100, 1,000, and 25,000 websites as ranked on Quantcast's top 1 million websites in the United States in July 2015. We collect data using two processes: 1) a

shallow automated crawl of the top 100, 1,000, and 25,000 sites, which consists of visiting only the homepage of the domain obtained from Quantcast's rankings, and 2) a deep automated crawl of the top 100 and 1,000 sites that consists of visiting the homepage and 2 randomly selected links from the homepage. After visiting the first link, the crawler returns to the homepage before selecting the second link. Both links are on the same domain as the homepage.

**The Crawler.** The crawler is OpenWPM, a flexible and scalable platform written in Python [31]. This crawler offers features such as collecting HTTP cookies, Flash cookies, HTML5 local storage objects, and the ability to perform deep crawls by visiting links. OpenWPM allows the crawl to be run in either Firefox or Chrome. It can be run with or without add-ons.

We run all crawls using a Firefox version 39 browser with no add-ons, with Flash turned on, and in headless mode. We collect information from each crawled domain visit: HTTP cookies, HTML5 local storage objects, Flash cookies, and HTTP requests and responses (including headers). We run each crawl four times and report the average for each tracking method.

**Shallow Automated Crawl.** We run shallow crawls with a clean browser instance cleared of all tracking data. The crawler visits each URL homepage, waits for the page to load, and then dumps all tracking data obtained from that URL into a database. The crawler then closes that browser tab, opens a new tab, then continues this process with the next URL on the Quantcast list.

**Deep Automated Crawl.** We run deep crawls with a clean browser instance cleared of all tracking data. The crawler visits each URL homepage and waits for the page to load. It then randomly selects a link on the homepage and visits that page. After the linked page finishes loading, the crawler goes back to the previous page and visits a second randomly selected link. After the second link finishes loading, the crawler dumps all tracking data obtained from those three URLs into a database. The crawler then closes that browser tab, opens a new tab, then continues this process with the next URL on the Quantcast list.

## Results

Top 100 Websites - Shallow Crawl

In our shallow crawl, we detected cookies on 99 of the top 100 websites, in comparison with all 100 in October 2012. In total, we detected 6,280 HTTP cookies for the top 100 websites, compared to 3,152 in October 2012. In 2015, with our shallow crawl, we found 3 websites that placed 300 or more cookies.

Figure 1 shows the distribution of cookies for the top 100 sites. The x-axis is the number of cookies, and the y-axis is the number of sites.

A significant number of top sites used Flash cookies, but the biggest increases are in the use of HTML5. In 2012, we found that 34 sites use HTML5. In this investigation, 76 sites used HTML5 when we investigated three links on the site. Also, many "keys" are included in HTML5 cookies. In our shallow crawl, we detected more than 800 keys in HTML5 storage.

Most HTTP cookies—83% of them—came from a third-party host. We detected 275 third-party hosts among the third-party cookies. This means that Internet tracking remains diffuse. A user who browses the most popular websites must vet dozens, even hundreds of policies to understand the state of data collection online.

At the same time, one player has an outsized ability to track online. Google Analytics had cookies on 15 sites; Google's ad tracking network, doubleclick.net, had cookies on 26 sites; youtube.com, also owned by Google, had cookies on 8 sites. Overall, Google had a presence on 85 of the top 100 websites.

Facebook had a presence on 20 of the top 100 websites.

The most frequently appearing cookie keys for the top 100 sites in our shallow crawl were: uid, _ga, __qca, i, __uuid.

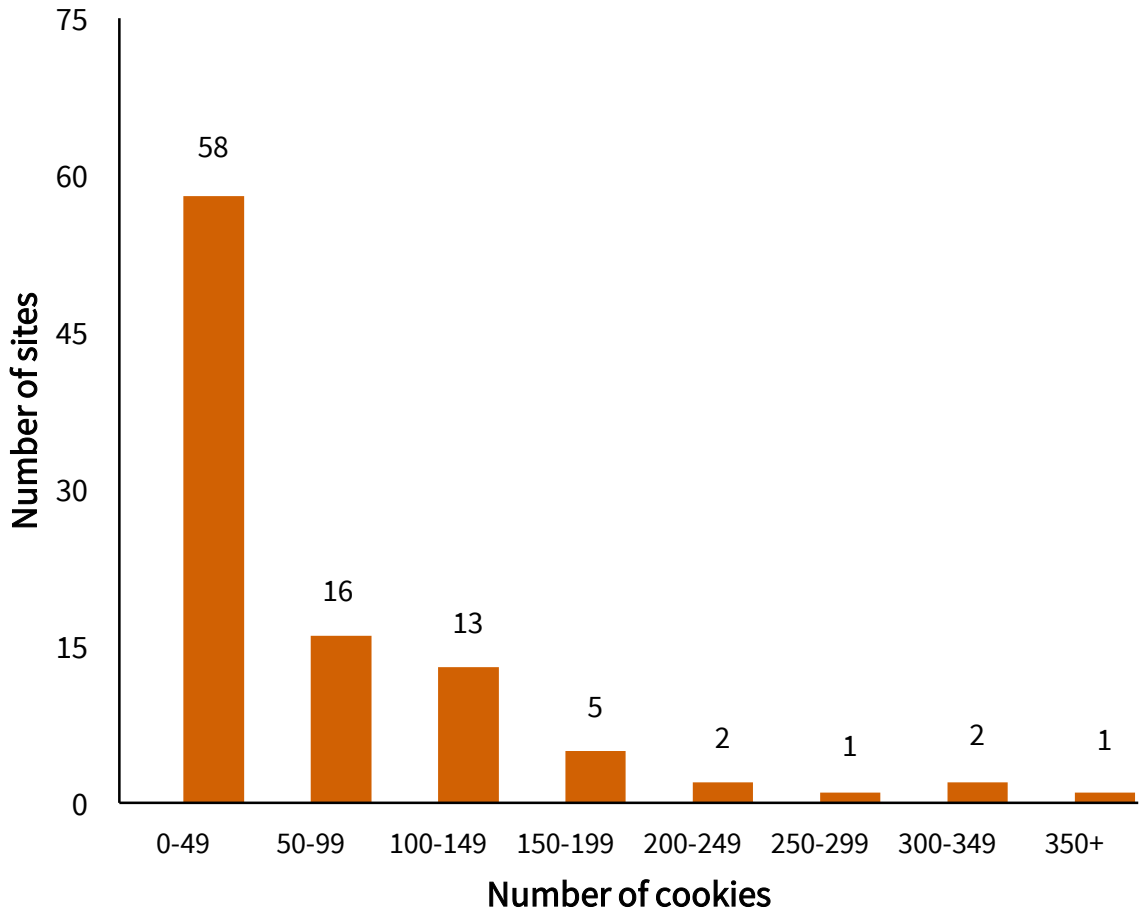

**Figure 1. Histogram showing number of HTTP cookies (horizontal axis) found on the top 100 websites using shallow crawl. Vertical axis is the number of websites with a given number of cookies.**

| Number of cookies | Number of sites |
|---|---|
| 0-49 | 58 |
| 50-99 | 16 |
| 100-149 | 13 |
| 150-199 | 5 |
| 200-249 | 2 |
| 250-299 | 1 |
| 300-349 | 2 |
| 350+ | 1 |

### Top 100 Shallow Flash Cookies and HTML5 Local Storage

Figure 2 shows an increase in the number of Flash cookies from 2012 to 2015 on the 100 most popular web pages using shallow crawl. We tracked 877 HTML5 storage keys for these same sites.

| Description | 2012 | 2015 | Trend |
|---|---|---|---|
| Number of Flash Cookies | 7 | 14 | up |
| Number of HTML5 Local Storage Keys | -- | 877 | -- |

**Figure 2. Historical comparison of Flash cookies and HTML5 storage from 2012 to 2015 appearing on the homepages of the top 100 websites using shallow crawl.**

Top 100 Websites—Deep Crawl

When we visited sites and made two clicks on the same domain, we detected cookies on all 100 top websites. In total, we detected 12,857 HTTP cookies for the top 100 websites, compared to 6,485 in October 2012. Figure 3 shows a summary of the key tracking metrics.

| Crawl Date | October 2012 | October 2015 | Trend |
|---|---|---|---|
| Do all popular sites have cookies | Yes | Yes | -- |
| Sites with 100 or more cookies | 21 | 45 | up |
| Sites with 150 or more cookies | 11 | 36 | up |
| Percentage of cookies set by a third party host | 84.7% | 93.5% | up |
| Number of third party hosts | 457 | 322 | down |
| Number of top websites with a Google presence | 74 | 92 | up |
| Number of sites with flash cookies | 11 | 10 | down |
| Number of sites with html5 storage | 38 | 76 | up |
| Number of sites without third party cookies | 5 | 6 | up |

**Figure 3. Key tracking metrics found in 2015 with comparisons to 2012.**

In 2015, our deep crawl found that 11 websites placed 300 or more cookies. Figure 4 shows a summary. Google Analytics had cookies on 52 of the top sites; doubleclick.net had cookies on 73 sites; YouTube had cookies on 19 sites. Overall, Google had a presence on 92 of the top 100 websites. Facebook had a presence on 57 of the top 100 websites.

Our observations about Flash cookies and HTML5 storage in the shallow crawl were also reflected in a crawl to three links on sites. Flash cookies grew modestly, but sites now use HTML5 to store many keys about site visitors.

The most frequently appearing cookie keys for the top 100 sites in our deep crawl were: _ga, uid, __utma, __utmz, id.

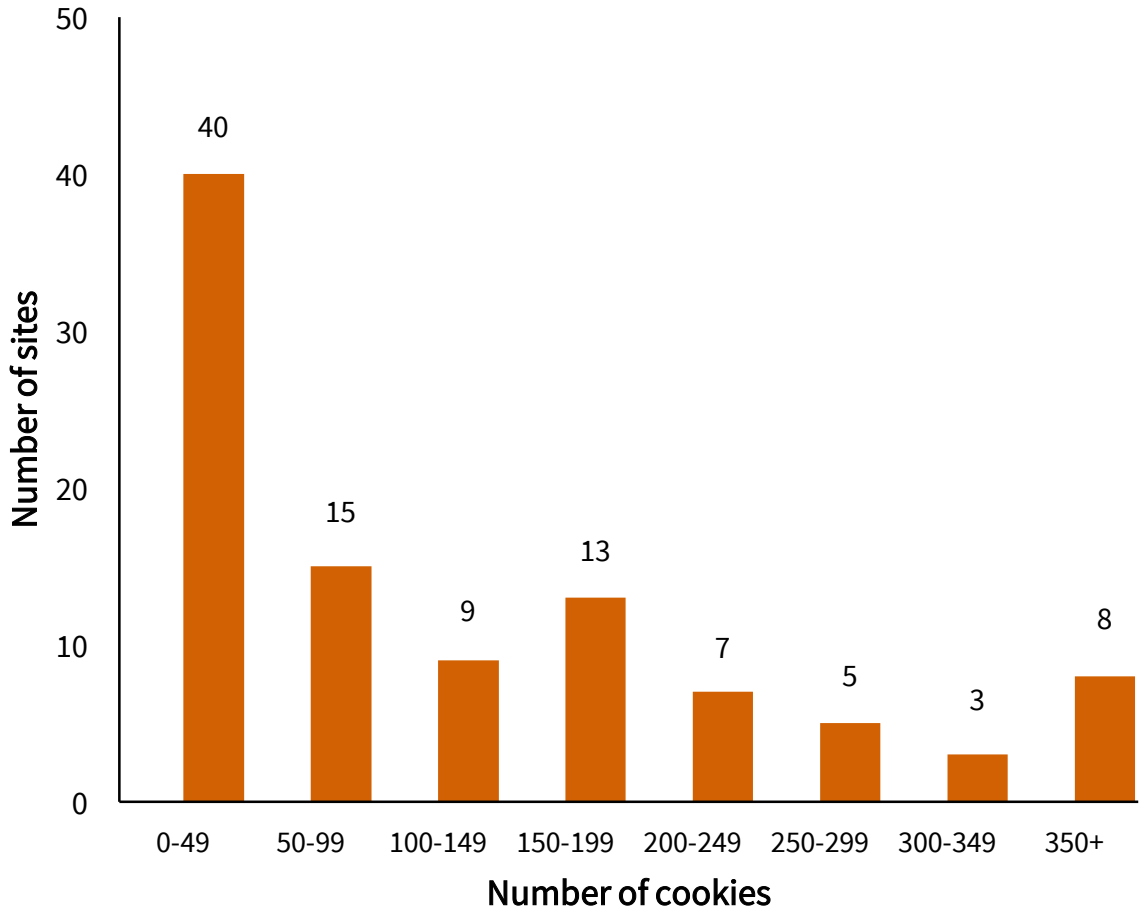


**Figure 4. Histogram showing number of HTTP cookies (horizontal axis) found on the top 100 websites using deep crawl. Vertical axis is the number of websites with a given number of cookies.**

| Number of cookies | Number of sites |
|---|---|
| 0-49 | 40 |
| 50-99 | 15 |
| 100-149 | 9 |
| 150-199 | 13 |
| 200-249 | 7 |
| 250-299 | 5 |
| 300-349 | 3 |
| 350+ | 8 |

## Top 1,000 Websites - Shallow Crawl

In 2015, with a shallow crawl, we detected cookies on 94% of the top 1,000 websites. In total, there were 80,821 HTTP cookies for the top 1,000 websites. Forty-six sites placed 300 or more cookies.

The most frequently appearing cookie keys for the top 1,000 sites in our shallow crawl were: _ga _utma, _utmz, _qca, uid.

Figure 6 shows the distribution of cookies for the top 1,000 sites. The x-axis is the number of cookies, and the y-axis is the number of sites. Most cookies—87% of them—were placed by a third-party host. We detected more than 797 third-party hosts among the third-party cookies. Google Analytics had cookies on 151 of the top sites; doubleclick.net had cookies on 212 sites; youtube.com had cookies on 65 sites. Overall, Google had a presence on 844 of the top websites.

Facebook had a presence on 182 of the top websites.

The most frequently appearing cookie keys for the top 1,000 sites in our shallow crawl were: _ga, __utma, __utmz, __qca, uid.

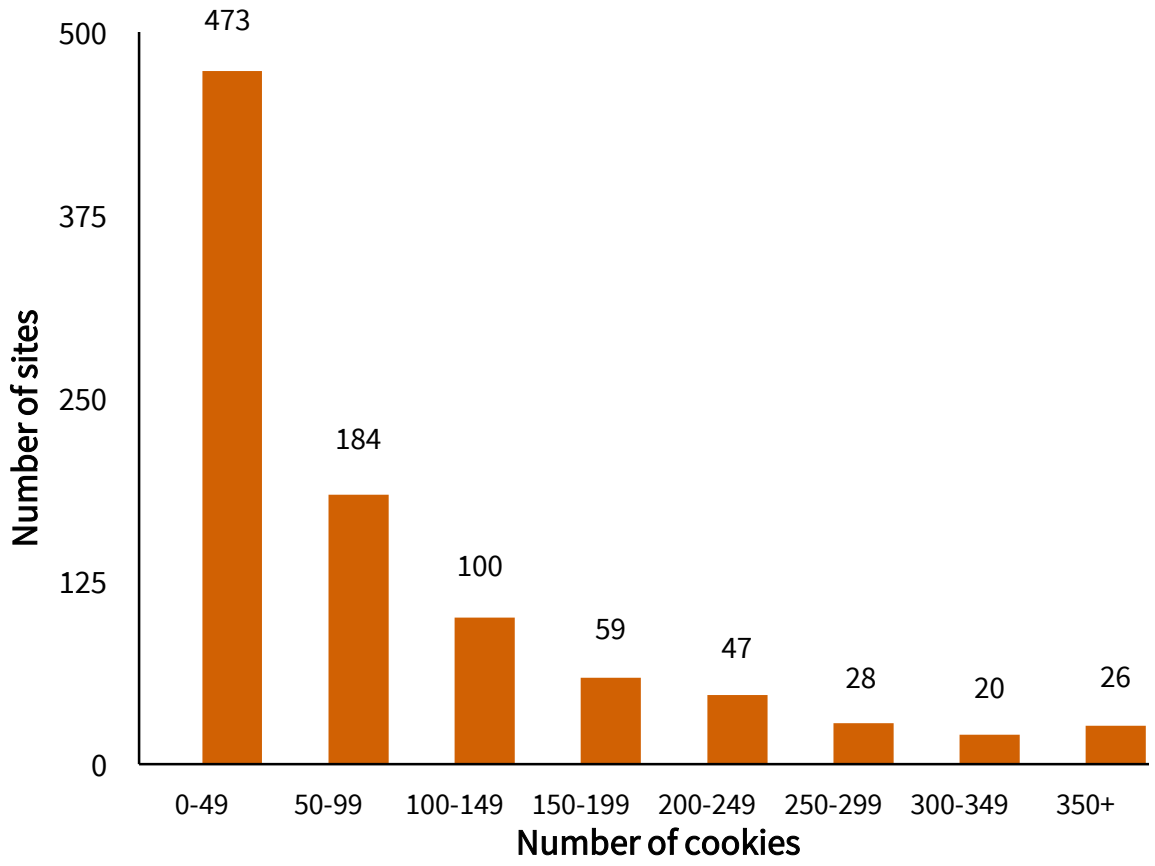

**Figure 6. Histogram showing number of HTTP cookies (horizontal axis) found on the top 1,000 websites using shallow crawl. Vertical axis is the number of websites with a given number of cookies.**

| Number of cookies | Number of sites |
|---|---|
| 0-49 | 473 |
| 50-99 | 184 |
| 100-149 | 100 |
| 150-199 | 59 |
| 200-249 | 47 |
| 250-299 | 28 |
| 300-349 | 20 |
| 350+ | 26 |

**Top 1,000 Websites - Deep Crawl**

In 2015, with a deep crawl, we detected cookies on 95% of the top 1,000 websites. In total, there were 134,769 HTTP cookies for the top 1,000 websites, compared to 65,381 in 2012. One hundred and thirty sites placed 300 or more cookies.

The most frequently appearing cookie keys for the top 1,000 sites in our deep crawl were: _ga, _utma, _utmz, optimizelySegments, optimizeltEndUserID.

Figure 8 shows the distribution of cookies for the top 100 sites. The x-axis is the number of cookies, and the y-axis is the number of sites. Most cookies—92% of them—were placed by a third-party host. We detected more than 880 third-party hosts among the third-party cookies.

Google Analytics had cookies on 581 of the top sites; doubleclick.net had cookies on 754 sites; youtube had cookies on 121. Overall, Google had a presence on 923 of the top websites.

Facebook had a presence on 548 of the top websites.

The most frequently appearing cookie keys for the top 1,000 sites in our deep crawl were: ga, utma, id, utmz, and optimizeitEndUserID.

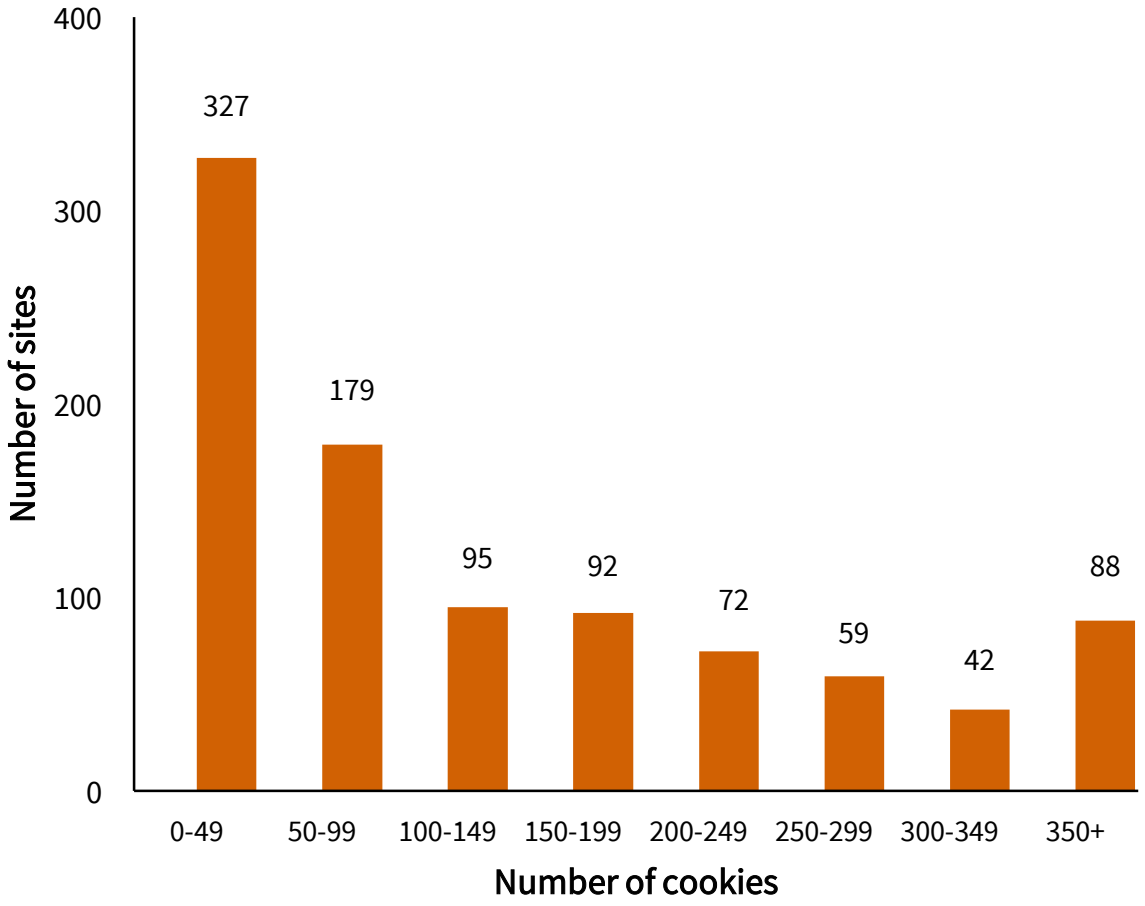


**Figure 8. Histogram showing number of HTTP cookies (horizontal axis) found on the top 1,000 websites using deep crawl. Vertical axis is the number of websites with a given number of cookies.**

| Number of cookies | Number of sites |
|---|---|
| **0-49** | 327 |
| **50-99** | 179 |
| **100-149** | 95 |
| **150-199** | 92 |
| **200-249** | 72 |
| **250-299** | 59 |
| **300-349** | 42 |
| **350+** | 88 |

### Top 1,000 Deep Flash Cookies and HTML5 Local Storage

Figure 9 shows an increase in the number of Flash cookies from 2012 to 2015 found on the 1,000 most popular web pages using deep crawl. We tracked 6,309 HTML5 storage keys for these same sites.

| Description | 2012 | 2015 | Trend |
|---|---|---|---|
| Number of Flash Cookies | 181 | 211 | up |
| Number of HTML5 Local Storage Keys | -- | 6309 | -- |

**Figure 9. Historical comparison of Flash cookies and HTML5 storage from 2012 to 2015 appearing on the homepages of the top 1,000 websites using deep crawl.**

Top 25,000 Websites—Shallow Crawl

We detected HTTP cookies on 81% of the top 25,000 websites. In total, we detected 1,065,076 HTTP cookies on the top 25,000 websites, compared to 476,492 in October 2012. In 2015, with our shallow crawl, we found 568 sites placing 300 or more cookies.

Figure 10 shows the distribution of cookies for the top 25,000 sites. The x-axis is the number of cookies, and the y-axis is the number of sites. Most cookies—87% of them—come from a third-party host. We detected more than 8,839 third-party hosts among the third-party cookies. Google Analytics had cookies on 11,521 of the top sites; doubleclick.net had cookies on 5,883; YouTube had cookies on 1,453. Overall, Google had a presence on 18,375 of the top 25,000 websites.

Facebook had a presence on 2,123 of the top websites.

The most frequently appearing cookie keys for the top 25,000 sites in our shallow crawl were:__utma,__utmz,_ga,__utmb,__gads,__qca.

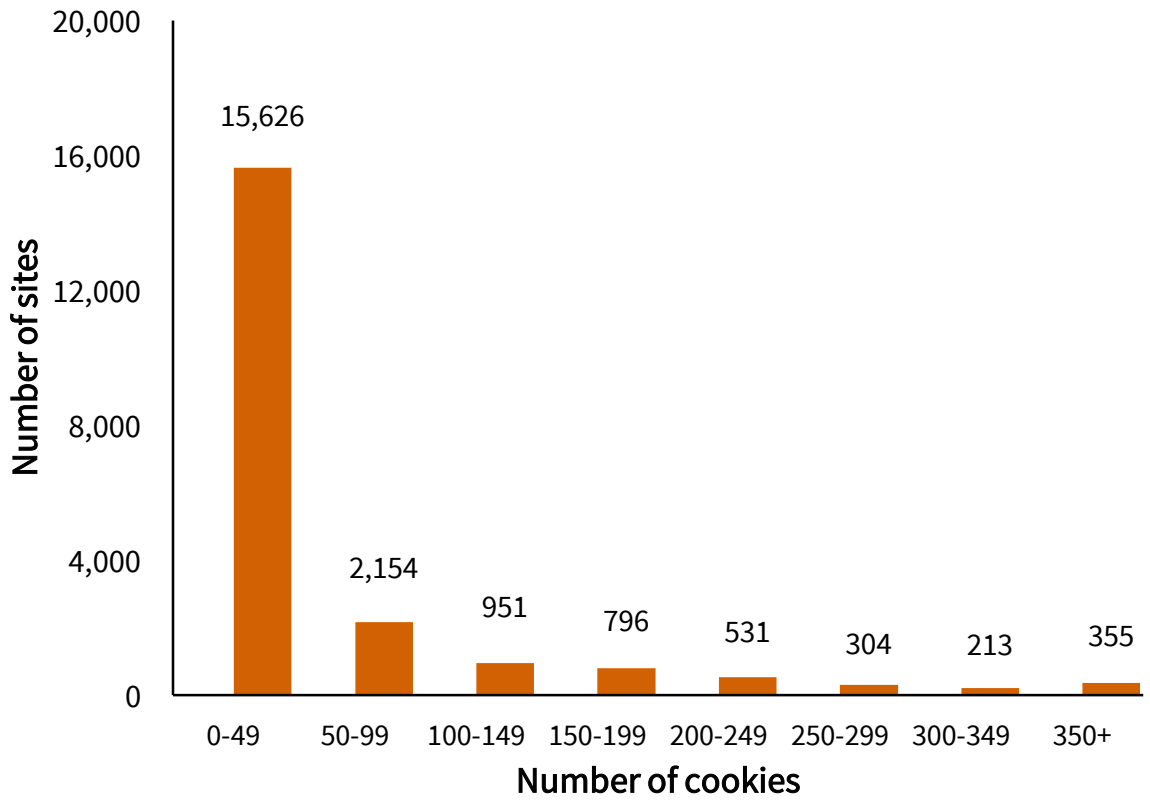


**Figure 10. Histogram showing number of HTTP cookies (horizontal axis) found on the top 25,000 websites using shallow crawl. Vertical axis is the number of websites with a given number of cookies.**

| Number of cookies | Number of sites |
|---|---|
| **0-49** | 15,626 |
| **50-99** | 2,154 |
| **100-149** | 951 |
| **150-199** | 796 |
| **200-249** | 531 |
| **250-299** | 304 |
| **300-349** | 213 |
| **350+** | 355 |

**Top 25,000 Shallow Flash Cookies and HTML5 Local Storage**

Figure 11 shows an increase in the number of Flash cookies from 2012 to 2015 on the 25,000 most popular web pages using shallow crawl. We tracked 48,949 HTML5 storage keys for these same sites.

| Description | 2012 | 2015 | Trend |
|---|---|---|---|
| Number of Flash Cookies | 454 | 1804 | up |
| Number of HTML5 Local Storage Keys | -- | 48949 | -- |

**Figure 11. Historical comparison of Flash cookies and HTML5 storage from 2012 to 2015 appearing on the homepages of the top 25,000 websites using shallow crawl.**

Figure 12 lists the names of the top trackers.

| Top Sites | Type | Top 10 Trackers | Number of Sites |
|---|---|---|---|
| 100 | shallow | quantserve.com | 33 |
| | | agkn.com | 28 |
| | | bluekai.com | 25 |
| | | doubleclick.net | 25 |
| | | rubiconproject.com | 24 |
| | | rlcdn.com | 24 |
| | | advertising.com | 23 |
| | | google.com | 22 |
| | | scorecardresearch.com | 21 |
| | | krxd.net | 20 |
| 1,000 | Shallow | google.com | 224 |
| | | doubleclick.net | 214 |
| | | quantserve.com | 208 |
| | | rlcdn.com | 187 |
| | | agkn.com | 185 |
| | | facebook.com | 185 |
| | | scorecardresearch.com | 182 |
| | | rubiconproject.com | 175 |
| | | bluekai.com | 169 |
| | | pubmatic.com | 166 |

**Figure 12. The top trackers found in the study and the number of distinct websites on which they were found.**

Table 2 shows a summary of the number of tracking technologies (HTTP cookies, Flash cookies and HTML5 cookies) returned by the top-level websites for the top 100, 1,000 and 25,000 domains we visited. It displays the sum per category as well as the percentage overall.

| Description | Date | Type | Sites Crawled | Total HTTP Cookies | First-Party HTTP Cookies | Third-Party HTTP Cookies | Sites Using Flash Cookies | Sites Using HTML5 Local Storage |
|---|---|---|---|---|---|---|---|---|
| **Top 100 Sites** | 2015-07-01 | shallow | 100 | 6,280 | 1,091 (17%) | 5,189 (83%) | 5 (5%) | 63 (63%) |
| **Top 100 Sites** | 2015-07-01 | deep | 100 | 1,2857 | 1,265 (10%) | 11,592 (90%) | 10 (10%) | 76 (76%) |
| **Top 1,000 Sites** | 2015-07-01 | shallow | 1,000 | 80,821 | 10,374 (13%) | 70,447 (87%) | 39 (4%) | 613 (61%) |
| **Top 1,000 Sites** | 2015-07-01 | deep | 1,000 | 134,769 | 10,871 (8%) | 123,898 (92%) | 62 (6%) | 649 (65%) |
| **Top 25,000 Sites** | 2015-07-01 | shallow | 25,000 | 1,065,076 | 135,767 (13%) | 929,309 (87%) | 585 (2%) | 8,688 (35%) |

**Table 2. Overall summary of results for shallow and deep crawls for the top 100, 1,000 and 25,000 websites.**

**Limitations of crawler methods.** For the October 2015 report, the crawler did not login to any sites, nor bypass any modal dialogs, and therefore our data does not record how cookies changed based on additional information provided by users logging into third-party services or requesting further access to the main site. Additionally, as the crawler automated selection of URLs for deep crawls, we did not necessarily capture any retargeting based on a human action (e.g., adding items to a shopping cart). We limited deep crawls to HTML anchor tags found and did not follow links set by JavaScript. Additionally, we randomly selected from links obtained by the deep crawler, and we consequently did not take into account page layout and visual layout in the selection process. We ran the crawl using Firefox with no add-ons.

**Limitations of data collection methods.** The identification and classification of third- and first-party cookies is a complex task. Many tracking and advertising companies are owned by other sites that have different domain names. For example, DoubleClick is owned by Google. For consistency in categorizing third-party cookies, the public suffix list was leveraged to combine suffixes consistent with previous work. We classified cookies from the top-level domain as first-party, while we classified cookies from a domain outside of the top-level domain third-party. This limited analysis of third-party domains to domains syntactically considered to be third parties. The analysis is not reflective of any underlying agreements or connections that may exist between multiple domains, through “DNS aliasing,” for instance, where a primary domain assigns a subdomain to a tracking company. Under such an arrangement, ordinary third-party cookies are instantiated in a first-party fashion. The ranking list used was Quantcast's top 1 million sites in the United States. This ranking may be different in other countries. Some websites on Quantcast’s top 1 million list don’t wish to be listed on the list and are marked as “Hidden profile”. We crawled top 100, 1000, and 25,000 excluding “hidden profiles”.

## Discussion

We found that users who merely visit the homepages of the top 100 most popular sites collect over 6,000 HTTP cookies in the process. —twice as many as we detected in 2012. If the user browses to just two more links, the number of HTTP cookies doubles. Third-party hosts set 83% of cookies. Just by visiting the homepage of popular sites, users receive cookies placed by 275 third-party hosts.

Some popular websites use many cookies. In just visiting the homepage of popular sites, we found 24 websites that placed 100 or more cookies, 6 websites that placed 200 or more cookies, and 3 websites that placed 300 or more cookies.

We also found that more sites are using HTML5 storage, which enables websites to store a greater amount of information about consumers.

By just visiting three links per site, we found that Google has tracking infrastructure on 92 of the top 100 most popular websites and on 923 of the top 1,000 websites. This means that Google's ability to track on popular websites is unparalleled and approaches the level of surveillance that only an ISP can achieve.

In comparison to 2012, tracking on the Web increased. There has been a marked increase in HTTP cookies and HTML5 storage usage. Cookie counts continued to increase, with larger amounts of third-party cookies in use. More than half of the top cookies ( _ga, __utma, __utmb, __utmz, optimizelyEndUserId) collect information on the pages visited by a user.

Google continues to be the single entity that can track individuals online more than any other company aside from a user's Internet Service Provider. Still, hundreds of third-party hosts also track users, and under the current self-regulatory regime [32], it is up to users to investigate these companies' privacy policies and decide whether to use the websites.

## References

1. Ayenson, M, Wambach D, , Soltani A, Good N, Hoofnagle, C. Flash Cookies and Privacy II: Now with HTML5 and ETag Respawning. July 29, 2011. http://ssrn.com/abstract=1898390.

2. Hoofnagle, Chris Jay and Good, Nathan, Web Privacy Census (June 1, 2012), available at: http://ssrn.com/abstract=2460547.

3. Electronic Privacy Information Center, Surfer Beware: Personal Privacy and the Internet. June 1997. https://epic.org/reports/surfer-beware.html.

4. Electronic Privacy Information Center, Surfer Beware II: Notice is Not Enough. June 1998. https://epic.org/reports/surfer-beware2.html.

5. Federal Trade Commission, Privacy Online: A Report to Congress. June 1998. http://www.ftc.gov/reports/privacy3/toc.shtm.

6. Electronic Privacy Information Center, Surfer Beware III: Privacy Policies without Privacy Protection. December 1999. https://epic.org/reports/surfer-beware3.html.

7. Federal Trade Commission, Privacy Online: Fair Information Practices In the Electronic Marketplace: A Report to Congress. May 2000. http://www.ftc.gov/reports/privacy2000/privacy2000.pdf.

8. Krishnamurthy B, Wills C. Privacy diffusion on the web: A longitudinal perspective, Proceedings of the 18th ACM international conference on World wide web. 2009. p. 541-550. http://portal.acm.org/citation.cfm?id=1526782.

9. Gomez J, Pinnick T, Soltani A. KnowPrivacy. June 1, 2009. http://www.knowprivacy.org/report/KnowPrivacy_Final_Report.pdf.

10. Soltani A, Canty S, Mayo Q, Thomas L, Hoofnagle C. Flash Cookies and Privacy. August 10, 2009. http://ssrn.com/abstract=1446862, accepted for publication at AAAI Spring Symposium on Intelligent Information Privacy Management, CodeX: The Stanford Center of Computers and Law.

11. Angwin J. The Web's New Gold Mine: Your Secrets, A Journal investigation finds that one of the fastest-growing businesses on the Internet is the business of spying on consumers. Wall Street Journal. July 30, 2010. http://online.wsj.com/article/SB10001424052748703940904575395073512989404.html.

12. Coates M. A Study of HTTPOnly and Secure Cookie Flags for the Top 1000 Websites. December 28, 2010. http://michael-coates.blogspot.com/2010/12/study-of-httponly-and-secure-cookie.html.

13. McDonald A, Cranor L. A Survey of the Use of Adobe Flash Local Shared Objects to Respawn HTTP Cookies. CMU-CyLab-11-001. 2011. http://www.casos.cs.cmu.edu/publications/papers/CMUCyLab11001.pdf.

14. Krishnamurthy B, Naryshkin K, Wills C. Privacy leakage vs. Protection measures: the growing disconnect. Presented at W2SP 2011: WEB 2.0 SECURITY AND PRIVACY. 2011. http://www.cs.wpi.edu/~cew/papers/w2sp11.pdf.

15. Mayer J. FourthParty, http://fourthparty.info/.

16. Mayer J. Tracking the Trackers: Early Results. July 12, 2011. http://cyberlaw.stanford.edu/node/6694.

17. Mayer J. Tracking the Trackers: To Catch a History Thief. July 19, 2011. http://cyberlaw.stanford.edu/node/6695.

18. Mayer J. Tracking the trackers: Where everybody knows your username. October 11, 2011. http://cyberlaw.stanford.edu/node/6740.

19. Mowery K, Shacham H. Pixel Perfect: Fingerprinting Canvas in HTML5. 2012. http://w2spconf.com/2012/papers/w2sp12-final4.pdf.

20. Wills C, Tatar C. Understanding What They Do With What They Know. 2012. https://dl.acm.org/citation.cfm?id=2381969.

21. Acar G. FPDetective: Dusting the Web for Fingerprinters. 2013. http://dl.acm.org/citation.cfm?id=2516674

22. Malandrino D, Petta A, Scarano V, Serra L, Spinelli R, Krishnamurthy B. Privacy Awareness about Information Leakage: Who Knows What About Me? 2013. http://www.di.unisa.it/~delmal/papers/UNISA-ISIS-082913TR.pdf.

23. Olejnik L, Castelluccia C, Janc A. Why Johnny Can't Browse in Peace: On the Uniqueness of Web Browsing History Patterns. 2012. http://hal.archives-ouvertes.fr/docs/00/74/78/41/PDF/johnny2hotpet-finalcam.pdf.

24. Mayer J, Mitchell J. Third-Party Web Tracking: Policy and Technology. 2012. https://www.ieee-security.org/TC/SP2012/papers/4681a413.pdf.

25. Krishnamurthy B. Privacy and Online Social Networks: Can Colorless Green Ideas Sleep Furiously? 2013. http://wiki.epfl.ch/edicpublic/documents/Candidacy%20exam/Privacy%20and%20Online%20Social%20Networks-%20Can%20colorless%20green%20ideas%20sleep%20furiously.pdf.

26. Mulazzani, M, Reschl P, Huber M, Leithner M, Schrittwieser S, Weippl E. Fast and Reliable Browser Identification with JavaScript Engine Fingerprinting. 2013. at http://obfuscationsymposium.org/wp-content/uploads/2014/02/gunes-position.pdf.

27. Nikiforakis N, Kapravelos A, Joosen W, Kruegel C, Piessens F, Vigna G. Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting. 2013. https://www.cs.ucsb.edu/~vigna/publications/2013_SP_cookieless.pdf.

28. Acar G. Obfuscation For and Against Device Fingerprinting Position Paper for Symposium on Obfuscation. February 15, 2014. http://obfuscationsymposium.org/wp-content/uploads/2014/02/gunes-position.pdf.

29. Reisman et al. Cookies That Give You Away: Evaluating the Surveillance Implications of Web Tracking. 2014. http://randomwalker.info/publications/cookie-surveillance.pdf.

30. Acar G, Eubank C, Englehardt S, Juarez M, Narayanan A, Diaz C. The Web Never Forgets: Persistent Tracking Mechanisms in the Wild. July 1, 2014. https://securehomes.esat.kuleuven.be/~gacar/persistent/the_web_never_forgets.pdf.

31. Englehardt S, et al. "OpenWPM: An Automated Platform for Web Privacy Measurement". Manuscript. March 2015.

32. Hoofnagle C, Urban J. Alan Westin's Privacy Homo Economicus. 49 Wake Forest Law Review 261. 2014. http://ssrn.com/abstract=2434800

---

## Authors

Ibrahim Altaweel is an undergraduate student studying computer science at University of California, Santa Cruz. He's a web security engineer at Good Research, and a privacy engineer at purrivacy.org. More information about Ibrahim Altaweel is available at his website at https://manip.io.

Dr. Nathan Good is Principal of Good Research and Faculty in UC Berkeley's Master of Data Science Program. He specializes in user experience research, modeling and investigating behavior where design overlaps with data. Past domains include systems for knowledge management, health care, sales support, consumer privacy, security and forensic tools, and recommender systems. Prior to Good Research, Nathan was at PARC, Yahoo and HP research labs. At Berkeley, he worked with TRUST and the Samuelson Law & Technology Clinic and was a member of the 2007 California Secretary of State Top-to-Bottom Review of Electronic Voting Systems. Nathan has published extensively on user experience studies, privacy, and security related topics and holds patents on software technology for multimedia systems and event analysis. Nathan's recent work on Privacy and Design was recognized for a best paper award at the Privacy Law Scholars Conference, and was featured in both IAPP and the Future of Privacy Forums top 6 Privacy Papers for Policy Makers. Nathan has a Phd in Information Science and a MS in Computer Science from the University of California at Berkeley.

Chris Jay Hoofnagle is adjunct full professor at the University of California, Berkeley, School of Information, and a faculty director of the Berkeley Center for Law & Technology. Hoofnagle teaches computer crime law, internet law, privacy law, and seminars on the Federal Trade Commission and on education technology. He is the author of Federal Trade Commission Privacy Law and Policy (Cambridge University Press 2006). Hoofnagle is of counsel to Gunderson Dettmer Stough Villeneuve Franklin & Hachigian, LLP, and a member of the

American Law Institute, the San Francisco Electronic Crimes Task Force, and Palantir's Council on Privacy and Civil Liberties.

This work was supported by TRUST, Team for Research in Ubiquitous Secure Technology, which receives support from the National Science Foundation (NSF award number CCF-0424422).

**Referring Editor:** Dierdre Mulligan

---

## Citation

Altaweel I, Good N, Hoofnagle C. Web Privacy Census. *Technology Science*. 2015121502. December 15, 2015. http://techscience.org/a/2015121502

---

## Data

Under review for data sharing classification.