Web Privacy Census: HTML5 Storage Takes the Spotlight As Flash Returns

Ibrahim Altaweel*‖, Jaime Cabrera†‖, Hen Su Choi‡‖, Katie Ho§‖, Nathan Good¶, Chris Hoofnagle¶
*Diablo Valley College, Pleasant Hill, CA 94523
Email: ialtaweel@berkeley.edu
†California State University Fullerton, Fullerton, CA 92831
Email: cabrera1@csu.fullerton.edu
‡University of California San Diego, La Jolla, California 92092
Email: henchoi@ucsd.edu
§Mount Holyoke College, South Hadley, Massachusetts 01075
Email: ho22k@mtholyoke.edu
¶University of California Berkeley, Berkeley, California 94720
‖ These authors contributed equally to this research.

***Abstract*—Building on previous studies, here we report on the state of internet tracking on the most popular web sites. We found a 38% increase in HTML5 local storage usage on 55 of the top 100 sites, as compared to 34 sites in 2012. We saw an unexpected increase in Flash cookies in 2014. Forty-one of the top 100 Global sites had flash cookies, with 25% originating from Chinese websites. Since the 2012 report, we have also noted a growth in alternative methods, besides HTTP cookies, that trackers have utilized to gather information from unsuspecting users on the internet. Here we compare data with the 2012 Web Privacy Census and discuss the patterns and trends we see surrounding the current state of web privacy.**

Keywords: web privacy census, HTML5 local storage, Flash cookies, HTTP cookies, privacy, online tracking

I. Introduction

Public policymakers are proposing measures to give consumers more privacy rights online. These measures are based upon the assumption that the web privacy landscape has become worse for consumers and that their online activities are tracked more pervasively now than they were in the past. In fact, since the 2012 Web Privacy Census, online advertising and metrics companies have developed more sophisticated ways to track and identify individuals online. This trend is substantiated in the academic literature, and in the popular press through an influential news series, "What They Know," by Wall Street Journal reporters. [2]

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, and direct regulation affect tracking. The first attempts of web measurement found relatively little tracking online in 1997–only 23 of the most popular websites were using cookies on their homepages. [3] But within a few years, tracking for network advertising was present on many websites, and by 2011, all of the most popular websites employed cookies. In 2014, users are still being tracked on sites by HTTP and Flash cookies. However, in 2013 and 2014, trackers have developed even newer techniques with more intrusive agendas to acquire users' personal information. In this report, we see a tendency of an increasing number of websites, as compared with data from 2012, to employ the use of HTML5 local storage objects.

Through this study, we seek to explore how many entities are tracking users online, what vectors are most popular in use for tracking users, and displacements in tracking practices.

In this paper, we will compare our data with the 2012 Web Privacy Census and provide an overview of prior related work on other tracking techniques. We outline our processes for gathering our data in the Methods section. Then, in the following two sections, we analyze the results from our crawler and consider important patterns and trends in our data. We conclude by reviewing the contrasting data between the 2012 and 2014 reports and discussing future work in the field.

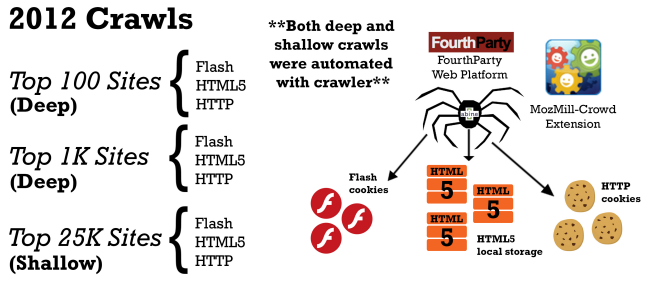

Fig. 1. For the 2012 report, shallow and deep crawls were automated for different sets of sites with the crawler.

A. Previous Work

In the 2012 report, the results of a crawl conducted on 5/17/12 were discussed and analyzed. Cookies were found on the top 100 U.S. most popular websites ranked by Quantcast. Two different crawls were conducted, as depicted by Figure 1. A shallow crawl where a test browser visited only the

homepage of a site, and a deep crawl where the browser visited six links on the same domain. Overall, it was found that flash cookie usage was dropping, HTML5 storage use was rising and at least one tracker was using HTML5 local storage to hold unique identifiers from third party cookies.

In the deep crawl of the top 100, cookies were detected on 100% of the sites. In total, 5795 HTTP cookies were found on the top 100 websites. 21 sites placed 100 or more cookies, including 6 that placed more than 150. In the distribution of the top 100 sites, 84% of the cookies were placed by a third party host. Google had a presence on 78 of the top websites. Only 22 of the top 100 US sites lacked some type of Google cookie.[1]

The top five third party trackers, the parties with a presence on the greatest number of sites, discovered on the top U.S. 100 sites were doubleclick.net, scorecardresearch.com, adnxs.com, quantserve.com, and atdmt.com. The top five trackers that placed the most cookies were bluekai.com, rubiconproject.com, pubmatic.com, doubleclick.net, and adnxs.com. The most frequently appearing cookie keys were utmb, utma, utmc, utmz, and pudm_AAAA. These keys were commonly associated with unique user tracking.

We found 23 Flash cookies on the top 100 sites compared to the 100 found in 2011. These Flash cookies appeared on 13 sites compared to 37 sites found in 2011. Additionally, 34 of the top 100 sites were using HTML5 local storage, double what we had seen in 2011. We did not collect data on HTML5 local storage in 2009.

In the deep crawl of the top 1000 sites, we detected cookies on 97.4% of the sites. In total, we detected 63,087 HTTP cookies for the top U.S. 1000 sites. 191 sites placed 100 or more cookies, including 117 that placed more than 150. Most cookies— 87% of them —were placed by a third party host. We detected over 2089 third party hosts among the 54,453 third party cookies. Google had a presence on 712 of the top US 1,000 sites. Only 285 lacked some type of Google cookie.[1]

The top five third party trackers on the 1000 sites were doubleclick.net, scorecardresearch.com, adnxs.com, quantserve.com, and atdmt.com. The top five trackers that set the most cookies on the 1000 sites were bluekai, rubiconproject.com, pubmatic.com, doubleclick.net, and adnxs.com. The most frequently appearing cookie keys were: utmb, utma, utmc, utmz, and pudm_AAAA. Many of these keys are commonly associated with unique user tracking.

176 Flash cookies were found on the top 1000 sites in a deep crawl. Those Flash cookies appeared on 110 of the top 1000 sites, while 311 sites were found to be using HTML5 local storage.

For the top 25,000 U.S. websites, a shallow crawl was conducted, hitting only the home page for each domain in the list and counting the cookies received. 87% of the top 25,000 websites had cookies. In total, 442,055 HTTP cookies were detected on the top 25,000 websites. 730 sites placed 100 or more cookies, including 133 that placed more than 150. In the distribution of cookies for the top 25,000 sites, 76% of them were placed by a third party host. More than 17,949 third party hosts were detected among the 334,011 third party cookies on the top 25,000 sites. Google had a presence on 8,993 of the top 25,000 websites. 15,596 lacked some type of Google cookie.[1]

The top five third party trackers on the 25,000 sites were doubleclick.net, quantserve.com, scorecardresearch.com, adnxs.com, and twitter.com. The top five trackers with most cookies were bluekai.com, doubleclick.net, adnxs.com, scorecardresearch.com, and casalemedia.com. The most frequently appearing cookie keys were: utmb, utma, utmc, utmz, and uid.

440 Flash cookies on the top 25,000 were found. These Flash cookies appeared on 344 sites. 2,416 of the top 25,000 sites were using HTML5 local storage.

B. Related Work

Although advertising companies have claimed that tracking is essential for the web economy to function, there is a significant risk in sharing users' private information with these entities, due to the possibility that such sensitive information will be leaked to other parties. Current solutions do not take into account the complete flow of data, the entities involved, and the volatility of data retention or leakage to external parties. Most of the time, privacy enforcement is done based on the entity that currently holds users' information, as opposed to different factors, such as which entity might end up with the data next and other unforeseen outcomes. [4]

HTTP cookies have continued to be the most used technique for third-party online tracking, but in recent years, a variety of more persistent mechanisms have also been introduced. [1] A considerable amount of work has been done in developing several alternatives to cookie tracking, including fingerprinting, respawning and cookie syncing.

Fingerprinting

1) Canvas Fingerprinting: Mowery and Shacham described how HTML5's canvas element could uniquely identify computers. This is accomplished through observing idiosyncrasies between the browser and operating system, which result in unique drawings rendered by canvas to be used as a fingerprint. [6]

When a user visits a webpage with the canvas fingerprinting script, the browser is instructed to draw text onto a canvas, which becomes an image. Then, this image is rendered to be used as a fingerprint. This type of tracking produces a unique fingerprint because each system draws a different image, with no notice to the user. Acar et al.'s 2014 study found Tor Browser's technique of sending a blank image as the only current software that was able to reliably combat canvas fingerprinting. [1]

At the time of this report, new research on this technique has just surfaced. There are ongoing studies doing more extensive research on the implications of this technique for

[1]We counted a Google presence as containing any cookies from the following domains: YouTube, DoubleClick, or Google.

users and exploring possible tools to counter this new type of tracking.

2) Browser & Device Fingerprinting: Olejnik et al. analyzed history-based user fingerprinting in their 2012 study. 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. [9]

In the 2013 study, Cookieless Monster: Exploring the Ecosystem of Web-Based Device Fingerprinting, Nikiforakis et al. surveyed over 800,000 users and conducted a 20-page crawl of Alexa's top 10,000 websites to detect fingerprinting. They found that 40 sites (0.4% of the Alexa top 10,000) are utilizing fingerprinting code. Additionally, they concluded that users who installed browser or user agent spoofing extensions perversely created a more unique fingerprint for themselves. The extensions aren't able to completely hide the browser's identity (ie. unable to spoof particular methods or properties), resulting in the user being even more recognizable. [8]

3) Javascript Engine Fingerprinting: Mulazzani et al. studied how spoofing a user agent string doesn't successfully hide the user's identity. They tested the underlying Javascript engine in multiple browsers and browser versions to find that they could, for the most part, reliably determine the user's browser without regard to the user agent at all. [7]

Cookie Syncing & Respawning

4) Respawning: In 2009, Soltani et al. demonstrated that popular websites were using Flash cookies to track users. Some advertisers adopted this technology because it allowed persistent tracking even where users had taken steps to avoid web profiling. They also demonstrated "respawning" on top sites with Flash technology. Respawning 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. [10]

Additionally, Acar et al. studied the method of respawning through the use of persistent cookies, known as evercookies. When a user visits a site with these persistent cookies, the site remembers a user's ID in different storage locations. Consequently, after a user clears their cookies, the site can still repopulate the cookie from the value saved in another storage location and recognize the user. [1]

5) Cookie Syncing: Acar et al. have also looked at the method of cookie syncing for tracking purposes. Through cookie syncing, pseudonymous IDs related to a particular user can be shared across different tracker domains. One way of passing this information from domain to domain is to pass the ID in a string through the URL. Websites have used this technique to share user information with one another to circumvent the fact that one domain cannot access another domain's cookies. The technique of cookie syncing allows trackers from different domains to compile their information and build up more comprehensive profiles on users. [1]

II. Methods

Data were collected on the top Global 100, 1000 and 25,000 websites from Alexa. These data were collected using two processes, outlined in Figure 2: 1) A shallow automated crawl of 100, 1000, and 25,000 sites, which consisted of visiting only the homepage of the domain obtained from Alexa's ranking, and 2) A shallow manual crawl of the top 100 and 400 sites from Alexa, to get the Flash cookies count and as a larger sample size to test the reliability of the crawler, respectively.

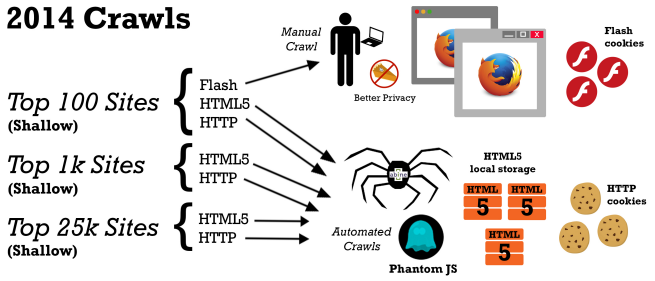

Fig. 2. For the 2014 report, only shallow crawls were automated for different sets of sites with the crawler. A manual crawl of 100 sites for Flash cookies was also conducted separately.

One of the objectives of the Web Privacy Census was to develop a crawling process that could be used to take regular samples of the tracking ecosystem over time. The list of domains were crawled by a distributed system built for this purpose. The crawler runs using PhantomJS and Webkit, an open source browser engine. PhantomJS is a Java Script tool that deploys headless web browsers to simulate user activity, as illustrated by Figure 3. However, headless web browsing can have some limitations, including not being able to collect Flash cookies. Therefore, we conducted multiple crawls and manual samples to help yield more accurate results.

The following information was collected from each crawled domain visit: http cookies, local storage objects, and http requests and responses (including headers).

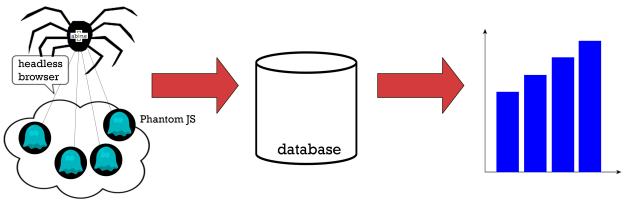

Fig. 3. Headless browser crawling sites and populating the database.

Shallow Automated Crawl

For the shallow crawl, each domain in the crawl list was visited with a fresh browser, clean data directories and the about:blank url loaded. The crawler navigated directly to the URL of the domain to be crawled and the browser and crawl settings to be used. Only the domain page was visited; therefore, no links were clicked or followed. After the homepage of the domain was loaded, all the data were stored and the crawler was cleaned out before it visited the next domain.

Shallow Manual Crawl

In order to test the reliability of the crawler, a manual crawl was performed on the top 400 sites from Alexa. For the manual crawl, each domain in the crawl list was visited with a fresh browser. The crawler visited only the homepage of each domain and collected the total number of HTTP cookies. The top 100 sites were checked for both Flash and HTTP cookies.

Limitations of crawler methods

For the 2014 report, the crawler still did not "log in" 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. Moreover, the headless browser did not have the capability to detect Flash cookies, so we did not record Flash cookie counts for the top 1000 or 25,000 sites.

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. Cookies from the top level domain were classified as first party, while cookies from a domain outside of the top level domain were classified as third party. Analysis of third party domains is therefore limited to domains that are syntactically considered to be third parties, and 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 would be instantiated in a first-party fashion. [5]

Additionally, our data from 2012 was based on the top U.S. sites according to Quantcast, while we conducted our crawls in 2014 with Alexa's list of top Global sites. There is possibility for variance between the lists.

III. Results & Discussion

Compared to the deep crawls of various sets of sites we ran in 2012, for the 2014 report, we only administered shallow crawls with larger sets of sites. Our shallow crawl of the top 25,000 sites revealed that 88% had HTTP cookies, and 35% had HTML5 storage objects. Furthermore, we found consistently, as illustrated in Figures 13, 14, and 15, that more popular sites set more cookies. As sites decrease in popularity and Alexa ranking, they steadily set fewer and fewer cookies.

In 2014, there was also a marked increase in flash cookies and HTML5 storage usage. HTML5 local storage objects allow developers more flexibility for storage and allow a much larger amount of information (5MB compared to 4KB in HTTP cookies) to be stored locally. [10] An increase of HTML5 storage is not directly connected with an increase in tracking, as the HTML5 storage object can hold any information that the developer needs to store locally. However, this storage can potentially contain information needed to consistently track users.

There was a surprising increase in Flash cookies count for the top Global 100 Alexa ranked sites in the 2014 report. The top sites with increased Flash cookie count in 2014 were primarily Chinese and news websites. Therefore, it is likely that the high count of Flash cookies on Chinese sites was not captured in previous years when the sample size was top U.S. sites. Overall, after past years of decline in numbers, it is expected that Flash cookie counts will drop in future years.

A. *Top 100 Sites*

We detected cookies on 95% of the top Global 100 websites, in comparison with 100% in 2012. In total, we found 1158 HTTP cookies with our shallow crawl for the top 100 websites. In 2012, using a deep crawl, we found that 21 sites placed 100 or more cookies, including 6 that placed more than 150. However in 2014, with our shallow crawl, we found no sites that placed 100 or more cookies.

The top five third party trackers, with a presence on the greatest number of sites, on the top 100 sites in 2014 were doubleclick.net, scorecardresearch.com, mmstat.com, adnxs.com, and yahoo.com. The top five third party trackers that set the most cookies on the top 100 sites were scorecardresearch.com, godaddy.com, go.com, doubleclick.net, and rubiconproject.com. Figures 4 and 5 identify these top five third party trackers.

Third Party Tracker	Number of Sites
.doubleclick.net	23
.scorecardresearch.com	16
.mmstat.com	9
.adnxs.com	7
.yahoo.com	6

Fig. 4. The top five trackers with the biggest presence on the top 100 sites.

The most frequently appearing cookie keys for the top 100 sites in our shallow crawl were: utma, utmb, utmc, utmz, and PREF. These keys are commonly associated with unique user tracking and Google Analytics. For instance, utma is used

Third Party Tracker	Total Number of Cookies Served
.scorecardresearch.com	32
.godaddy.com	30
.go.com	26
.doubleclick.net	25
.rubiconproject.com	24

Fig. 5. The top five trackers with the number of cookies they set on the top 100 sites.

by Google for identifying unique visitors, and PREF is used by Google to remember a user's preferences and preferred language.

We found 66 Flash cookies, with our shallow manual crawl, on the top 100 sites compared to the 23 cookies found with a deep crawl found in 2012. These Flash cookies appeared on 41 sites compared to 13 sites found in 2012. This increase is particularly surprising for multiple reasons. According to data we presented in the 2012 report and represented in Figures 6 and 7, Flash cookies have been on a steady decline over the past few years. Moreover, the 66 cookies were counted from only visiting the homepage, while the 23 cookies from 2012 were tallied from visiting the homepage and six other links.

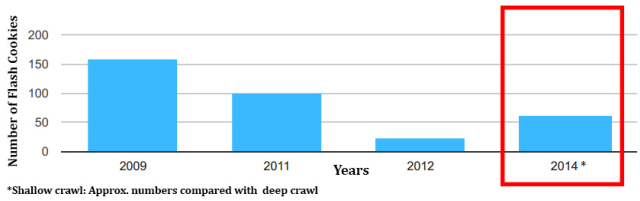

Fig. 6. The number of Flash cookies on the Top 100 sites over the past years.

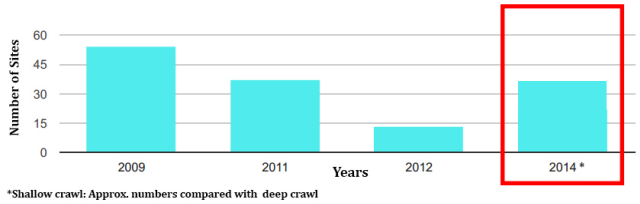

Fig. 7. The number of Top 100 sites with Flash cookies over the past years.

As of 2014, 56 of the top 100 sites are using HTML5 local storage, which is a notable 38% increase from the 34 sites using HTML5 local storage in 2012.

B. *Top 1000 Sites*

In 2012, with a deep crawl, we detected cookies on 97.4% of the top Global 1000 websites. In total, there were 63,087 HTTP cookies for the top 1000 websites. 191 sites placed 100 or more cookies, including 117 that placed more than 150.

With our shallow crawl in 2014, we detected cookies on 91.3% of the top 1000 websites. In total, there were 17,160 HTTP cookies for the top 1000 websites. 14 sites placed 100 or more cookies, including 10 that placed more than 150.

The top five third party trackers, with a presence on the greatest number of sites, on the top 1000 sites in 2014 were doubleclick.net, scorecardresearch.com, google.com, yahoo.com, quantserve.com. The top five third party trackers on the top 1000 sites that set the most cookies were rubiconproject.com, scorecardresearch.com, doubleclick.net, bluekai.com, casalemedia.com. Figures 8 and 9 identify these top five third party trackers.

The most frequently appearing cookie keys for the top 1000 sites in our shallow crawl were: utma, utmb, utmc, utmz, ID, NID. These keys are primarily used in Google Analytics, including ID and NID for advertising customization purposes.

Third Party Tracker	Number of Sites
.doubleclick.net	351
.scorecardresearch.com	211
.google.com	111
.yahoo.com	86
.quantserve.com	86

Fig. 8. The top five trackers with the biggest presence on the top 1k sites.

Third Party Tracker	Total Number of Cookies Served
.rubiconproject.com	425
.scorecardresearch.com	423
.doubleclick.net	379
.bluekai.com	333
.casalemedia.com	223

Fig. 9. The top five trackers with the number of cookies they set on the top 1k sites.

We found 176 Flash cookies on the top 1000 sites in 2012. These Flash cookies appeared on 110 sites. We did not collect data on Flash cookies for the top 1000 sites through our automated crawler in 2014.

505 of the top 1000 sites were using HTML5 local storage in 2014, also a 38% increase from 311 sites in 2012.

C. *Top 25,000 Sites*

For the top Global 25,000 websites in 2014, we performed a shallow crawl, hitting only the home page for each domain in the list and counting the cookies we received. The goal of this larger sample size was to get cookie counts for a wider range of sites to develop a more extensive understanding of trackers. We detected cookies on 88% of the top 25,000 websites. In total, we detected 355,524 HTTP cookies for the top 25,000 websites, compared with the cookie count from 2012, which detected 442,055 on the top 25,000 sites.

In 2012, 730 sites placed 100 or more cookies, including 133 that placed more than 150; while in 2014, 25 sites placed 100 or more cookies, including 23 that placed more than 150.

Third Party Tracker	Number of Sites
.doubleclick.net	8455
.google.com	4000
.scorecardresearch.com	3508
.twitter.com	2460
.yahoo.com	1757

Fig. 10. The top five trackers with the biggest presence on the top 25k sites.

Third Party Tracker	Total Number of Cookies Served
.doubleclick.net	9766
.scorecardresearch.com	7037
.rubiconproject.com	6384
.bluekai.com	6038
.pubmatic.com	5102

Fig. 11. The top five trackers with the number of cookies they set on the top 25k sites.

The top five third party trackers, with a presence on the greatest number of sites, on the top 25,000 sites in 2014 were doubleclick.net, google.com, scorecardresearch.com, twitter.com, yahoo.com. The top five third party trackers on the top 25,000 sites that set the most cookies were doubleclick.net, scorecardresearch.com, rubiconproject.com, bluekai.com, and pubmatic.com. Figures 10 and 11 identify these top five third party trackers.

The most frequently appearing cookie keys in the top 25,000 sites in our shallow crawl were: utma, utmb, utmc, utmz, ga, cfduid, gads. These keys are mainly associated with Google and its various services, including advertising.

We found 440 Flash cookies on the top 25,000 in 2012. These Flash cookies appeared on 344 sites. We did not collect data on Flash cookies for the top 25,000 sites through our automated crawler in 2014.

We expected an increase in HTML5 local storage objects on sites from our 2012 study, but the volume of the difference between the 2012 and 2014 reported numbers, a 72% increase, is surprising. In 2014, as depicted in Figure 8, we found 8,758 of the top 25,000 sites were using HTML5 local storage, compared with 2,416 sites in 2012.

IV. Conclusion

HTTP cookies, Flash cookies, and HTML5 local storage objects are undoubtedly popular mechanisms to uniquely identify and track users online. However, they are all methods that leave a trace, so if a user wanted to check their cookies or HTML5 local storage objects, the values would be easily accessible.

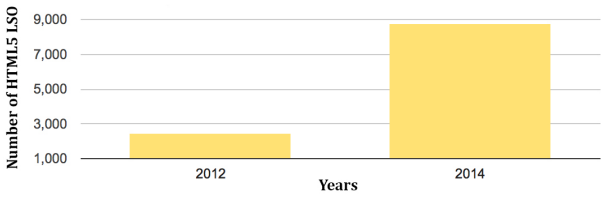

Fig. 12. The number of HTML5 local storage objects found on the Top 25,000 sites, comparing 2012 and 2014.

Unfortunately, if trackers shift away from these mechanisms that simply place files on user computers, towards fingerprinting, which is a type of tracking that can occur without leaving a trace, users will be much more limited in self-help against those who seek to obtain their information. This report reviewed related work on current fingerprinting techniques, but primarily focused on the more traditional tracking mechanisms that were measured in the 2012 report.

The 2014 Web Privacy Census is part of an ongoing study about web tracking. Future reports will continue to look at trends occurring over time and provide updated measures used by web trackers and technologies described in the cumulative reports.

In subsequent years, it is expected that Flash cookies will decrease in use, due to this spike in Flash cookie count being considered an anomaly. We predict that HTTP cookies will decline in popularity as well, in light of trackers' new use of fingerprinting techniques. It can be speculated that trackers and sites will employ the use of fingerprinting more broadly to track users, thus shifting away from cookies, which appeared to be the most common tracking technique in previous years.

Users are constantly advised to self-regulate the personal information they post online and learn more about the browser privacy plugins and tools available to them. However, tracking entities and cyber criminals are continuously inventing new ways to circumvent these practices. Consequently, the issue of internet privacy has become a technological arms race, with different parties working to conceal and reveal personal information from the unsuspecting public online.

Acknowledgements

This work was supported in part by TRUST, Team for Research in Ubiquitous Secure Technology, which receives support from the National Science Foundation (NSF award number CCF-0424422). Special thanks to Chris Hoofnagle, Nathan Good, Aimee Tabor, and the TRUST staff for their guidance and contributions to our research.

References

[1] Gunes Acar, Christian Eubank, Steven Englehardt, Marc Juarez, Arvind Narayanan, and Claudia Diaz. The web never forgets: Persistent tracking mechanisms in the wild.

[2] Julia Angwin. The web's new gold mine: Your secrets. *Wall Street Journal*, 2010.

[3] Electronic Privacy Information Center. Surfer beware: Personal privacy and the internet. https://epic.org/reports/surfer-beware.html. Accessed: 1997-06-01.

[4] Balachander Krishnamurthy. Privacy and online social networks: can colorless green ideas sleep furiously? *Security & Privacy, IEEE*, 11(3):14–20, 2013.

[5] Balachander Krishnamurthy and Craig Wills. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th international conference on World wide web*, pages 541–550. ACM, 2009.

[6] Keaton Mowery and Hovav Shacham. Pixel perfect: Fingerprinting canvas in html5. *Proceedings of W2SP*, 2012.

[7] Martin Mulazzani, Philipp Reschl, Markus Huber, Manuel Leithner, Sebastian Schrittwieser, Edgar Weippl, and FH Campus Wien. Fast and reliable browser identification with javascript engine fingerprinting. In *Web 2.0 Workshop on Security and Privacy (W2SP)*, volume 5, 2013.

[8] Nick Nikiforakis, Alexandros Kapravelos, Wouter Joosen, Christopher Kruegel, Frank Piessens, and Giovanni Vigna. Cookieless monster: Exploring the ecosystem of web-based device fingerprinting. In *Security and Privacy (SP), 2013 IEEE Symposium on*, pages 541–555. IEEE, 2013.

[9] Lukasz Olejnik, Claude Castelluccia, Artur Janc, et al. 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)*, 2012.

[10] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle. Flash cookies and privacy. In *AAAI Spring Symposium: Intelligent Information Privacy Management*, 2010.

APPENDIX

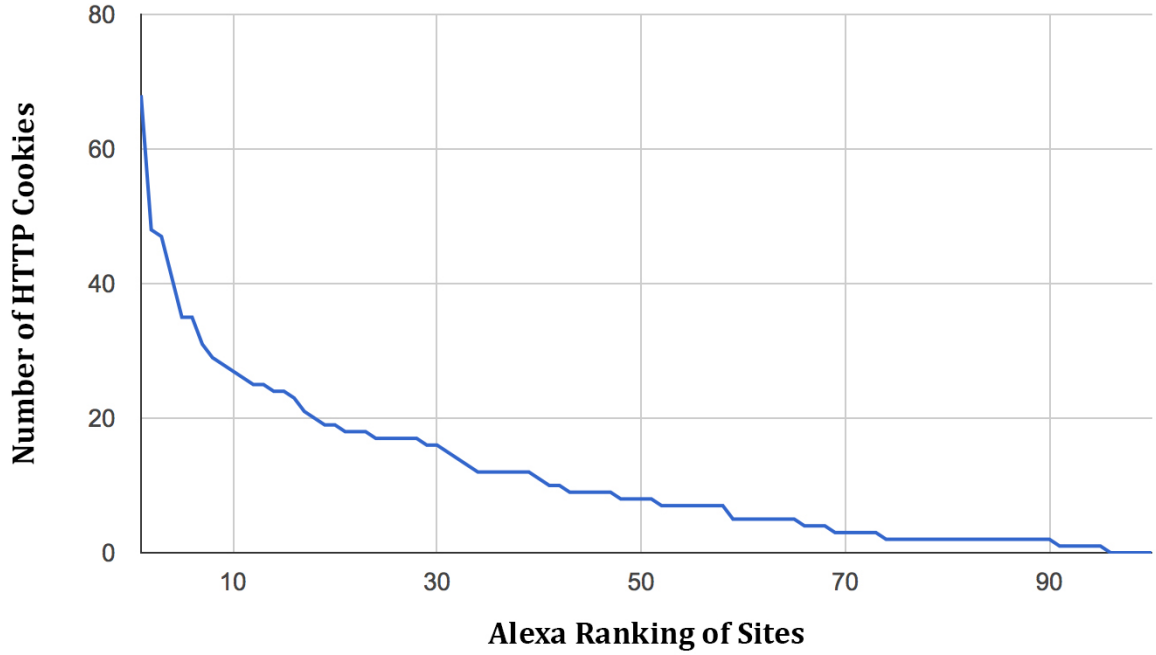

Fig. 13. The distribution of cookies for the top 100 sites. The y axis is the number of cookies, the x-axis is the sites ordered by the total number of cookies.

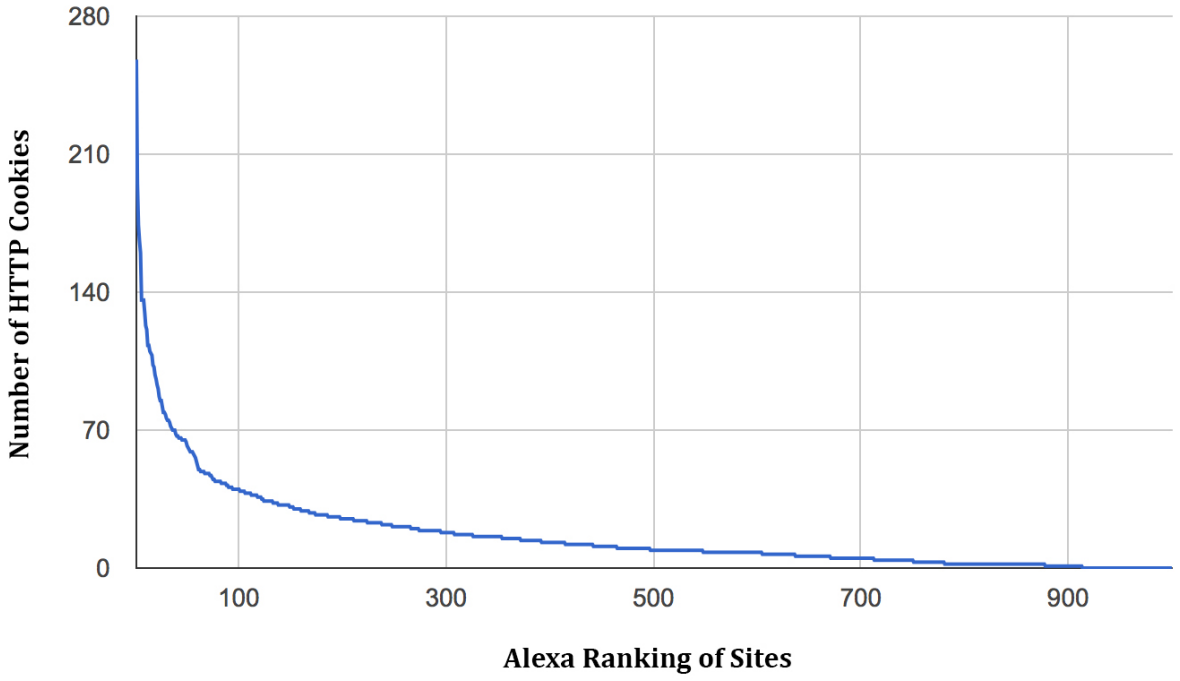

Fig. 14. The distribution of cookies for the top 1,000 sites. The y axis is the number of cookies, the x-axis is the sites ordered by the total number of cookies.

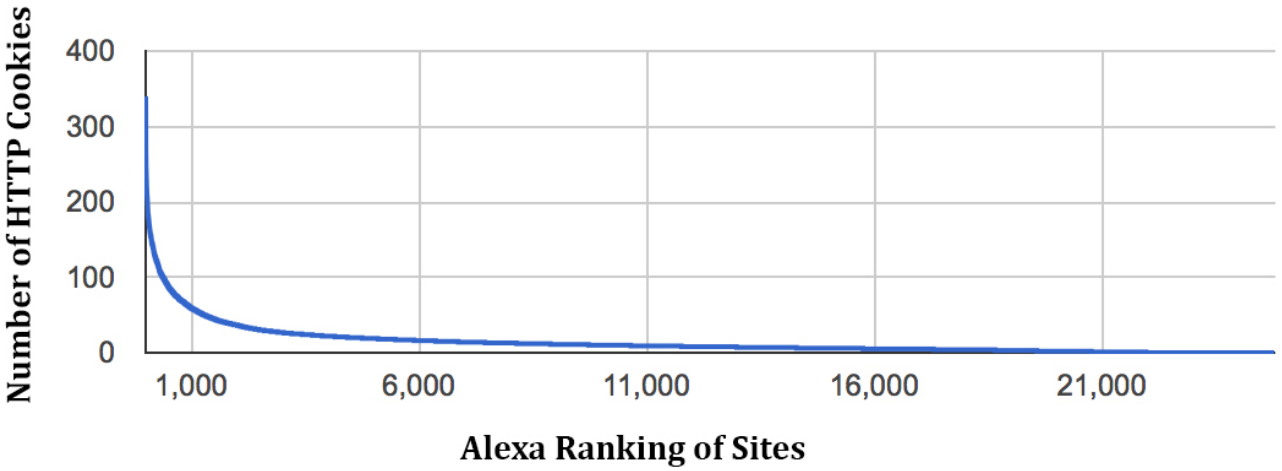

Fig. 15. The distribution of cookies for the top 25,000 sites. The y axis is the number of cookies, the x-axis is the sites ordered by the total number of cookies.