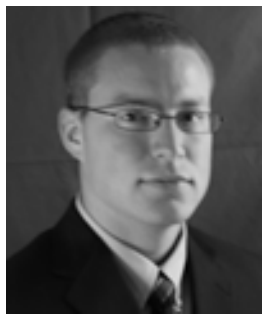
BNA

Privacy & Security Law Report®

Can Advertisers Learn That "No Means No"?




By Chris Jay Hoofnagle, Ashkan Soltani, Nathaniel Good, Dietrich J. Wambach, and Mika D. Ayenson

Chris Jay Hoofnagle, Lecturer in Residence, UC Berkeley Law. Ashkan Soltani, MIMS (Berkeley 2009), independent researcher and consultant focused on privacy, security, and behavioral economics. Nathaniel Good, Ph.D., Chief Scientist and Principal of Good Research. Dietrich J. Wambach is a senior at the University of Wyoming. Mika D. Ayenson is a junior at Worcester Polytechnic Institute.

Summary

Our research[1] examines two key issues in the online advertising debate: how can advertisers track users without their knowledge, and how they can override users' attempts to avoid tracking. In 2009, we

[1] This work was supported exclusively by TRUST (Team for Research in Ubiquitous Secure Technology), which receives support from the National Science Foundation (NSF award number CCF-0424422) and the following organizations: AFOSR (#FA9550-06-1-0244), BT, Cisco, ESCHER, HP, IBM, iCAST, Intel, Microsoft, ORNL, Pirelli, Qualcomm, Sun, Symantec, Telecom Italia, and United Technologies. We are grateful for the opportunities offered by the TRUST Research Experiences for Undergraduates program (REU), and to its program leader, Dr. Kristen Gates. The full version of this report is available at http://papers.ssrn.com/sol3/papers.cfm?

 ISSN 1538-3423

found that many popular websites were using Flash cookies, a technology that gave advertisers the ability to track users or back up ordinary cookies when the user deleted them. In our most recent paper, we find a decline in the use of Flash cookies, but observe that websites are using HTML5 and cache cookies as tracking mechanisms. The latter vector relies upon the user's browser cache to store tracking objects, and can persistently enumerate a user employing private browsing mode. We conclude by elucidating the privacy problems with these practices: poor notice, circumvention of user choice, the creation of platforms that enable websites to buy the information that users chose not to share with the site, and the hypocrisy of advertisers who bemoan ''paternalistic'' privacy rules while using technology to impress their own preferences upon users.

'No Means No.'

In a study of popular websites in 2009, we found widespread use of ''Flash cookies.''[2] Flash cookies, technically called ''local shared objects,'' are files used by Adobe Flash programs to store data on users' computers. Our 2009 paper documented that some advertisers adopted Flash cookies specifically because they were relatively unknown, more difficult for consumers to delete, and were more effective in tracking than HTTP cookies.[3] They could also be used to ''respawn'' (this is sometimes referred to as ''reinstantating'' or ''backing up'') ordinary HTTP cookies after a user had deleted them.

In 2011, Aleecia McDonald and Lorrie Faith Cranor of Carnegie Mellon found a dramatic decline in Flash cookie use.[4] They found that only 20 percent of top 100 websites used Flash cookies, and that only two sites respawned using Flash cookies. McDonald et al. 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.

But McDonald et al. used different methods than we did in 2009. We thus replicated our 2009 methods to benchmark the state of Flash cookie tracking. We visited the top 100 websites, collecting HTTP cookies, HTML5 cookies, and Flash cookies. We found that fewer websites were employing Flash cookies, but several other developments proved to be more important: first, one website that recently settled a suit for respawning cookies continued to use the technique with both Flash and cache cookies (specifically, ETags).[5] Second, we detected over 600 third party hosts tracking users on popular sites. Third, consistent with the findings of other researchers, there is strong concentration in third-party tracking online.

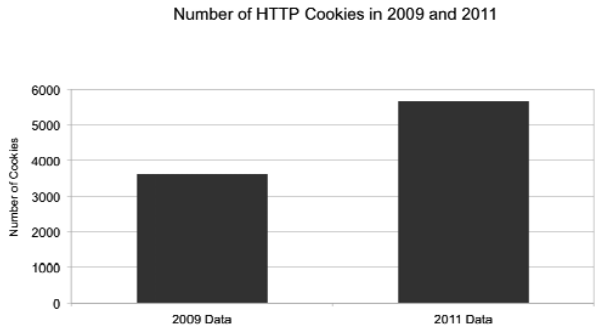
Since our 2009 paper, there have been a number of privacy developments with regard to Flash cookies. Flash cookies can now be controlled through browser privacy settings, and are managed in ''private browsing modes.'' Adobe itself criticized the practice of respawning[6] and the Network Advertising Initiative (NAI) felt compelled to issue a pusillanimous admonition on the use of Flash cookies.[7]

This focus on Flash, and the emergence of additional cache-based respawning and tracking shows that advertisers missed the point of our 2009 work. We tried to impress upon the industry that using technical methods to circumvent user choice and to mask tracking was unfair to users. The method of doing this—whether Flash, HTML5, or cache ETags, is irrelevant. Advertisers must learn that ''no means no.'' In other contexts, this principle has been lost on advertisers, and the industry's intransigence has resulted in promulgation of rules to protect consumer choice.[8]

Our Findings

HTTP Cookies

We first enumerated the prevalence of HTTP cookies on all top 100 websites. In total, we detected 5,675 HTTP cookies. This is dramatically higher than the 3,602 we detected in 2009 for the same category of sites (top 100 most popular sites). Twenty sites placed 100 or more cookies, including seven that placed more than 150.



Most cookies—4915 of them—were placed by a third party host. That is, a website other than the one the user is knowingly visiting or establishing a relationship with.

abstract_id=1898390 and more technical details are elucidated at http://ashkansoltani.org/docs/respawn_redux.html.

[2] Ashkan Soltani, Shannon Canty, Quentin Mayo, Lauren Thomas, and Chris Jay Hoofnagle, *Flash Cookies and Privacy*, Aug. 10, 2009, available at http://ssrn.com/abstract=1446862, *accepted for publication at* AAAI Spring Symposium on Intelligent Information Privacy Management 2010, CodeX: The Stanford Center of Computers and Law.

[3] HTTP cookies are what we normally refer to as ''browser cookies'' and enable websites to store information about a user, such as the contents of their shopping cart or more commonly, a unique tracking token.

[4] McDonald, A. M., & Cranor, L. F., *A Survey of the Use of Adobe Flash Local Shared Objects to Respawn HTTP Cookies*, CMU-CyLab-11-001 (2011), available at http://www.casos.cs.cmu.edu/publications/papers/CMUCyLab11001.pdf.

[5] ETags are tokens presented by a user's browser to a remote webserver in order to determine whether a given resource (such as an image) has changed since the last time it was fetched. It is typically used for version control.

[6] Adobe Systems Inc., Comments from Adobe Systems Incorporated – Privacy Roundtables Project No. P095416, Jan. 27, 2010, available at http://www.ftc.gov/os/comments/privacyroundtable/544506-00085.pdf (emphasis in original).

[7] Network Advertising Initiative, FAQs (n.d.), available at http://www.networkadvertising.org/managing/faqs.asp#question_19.

[8] See e.g. 16 CFR § 310.4(b)(ii) Abusive telemarketing acts or practices.

  PVLR ISSN 1538-3423

We detected over 600 third party hosts among the 4915 third party cookies. Google had cookies on 89 of the top 100 sites; the company's ad tracking network, doubleclick.net, had cookies on 77. Combined, Google has a presence on 97 of the top 100 websites. This includes popular government websites such as usps.com, irs.gov, and nih.gov.

Flash Cookies–Local Shared Objects

We found 100 Flash objects on the top 100 sites, down from the 281 we found in 2009. These Flash cookies appeared on 37 sites, down from the 54 sites we found in 2009.

Flash cookies can store many keys and values. MTV-.com had 8 flash cookies, one of which stored over 140 values. We found 454 key/value pairs in 100 Flash cookies detected.

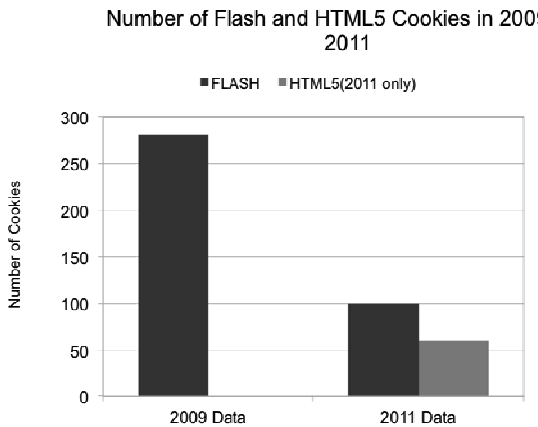
Two sites had shared values between Flash cookies and HTTP cookies: hulu.com and foxnews.com. The value was shared in HTML5 local storage as well.

HTML5 Storage

HTML5 local storage may soon become the tracking technology of choice among advertisers. Like Flash cookies, HTML5 storage is more persistent than HTTP cookies. HTTP cookies expire by default, and in order to make them persistent, developers must use a complex syntax and constantly update the expiration date. HTML5 data are persistent until affirmatively deleted by a web site or user. Storage size is important too. While Flash cookies have a default limit of 100KB, HTTP cookies store just 4KB, compared to 5Mb for HTML5 storage.[9]

Seventeen of the top 100 sites were using HTML5 local storage. These 17 sites had a total of 60 key/value pairs.

We found matching values among HTML5 local storage and HTTP cookies in several cases. In most of these cases, the matching value was with a third party service, such as meebo.com, kissanalytics.com, and polldaddy.com.



[9] Bruce Lawson & Remy Sharp, INTRODUCING HTML5 142-3 (New Riders 2011).

Table 1: Key Characteristics of HTTP Cookies, Flash Cookies, and HTML5 Storage

	HTTP Cookies	Flash cookies	HTML5 storage
Storage	4KB limit	100KB by default	5Mb by default
Expiration	Session by default	Permanent by default	Permanent by default
Location	In SQL file (Firefox)	Stored outside the browser	In SQL file (Firefox)
Access	Only by browser	By multiple browsers on same computer	Only by browser

Respawning

We found three respawning behaviors on two sites: hulu.com and foxnews.com.

In 2009, we reported that a QuantCast cookie was respawned on hulu.com. After our 2009 paper, QuantCast executives contacted authors Hoofnagle and Soltani almost immediately, and quickly acted to change the behavior of their service in order to prevent respawning.[10] We thus were surprised to find two new and different methods of cookie respawning on hulu-.com (completely unrelated to QuantCast).

First, hulu.com used Flash and HTML5 based respawning to reinstantiate a HTTP cookie with the key ''guid,'' mirroring a stored object with the key ''computerguid.''[11] Unlike the situation in 2009, where a third party respawned the cookies, this use of Flash/HTML5 storage was enabled in-house by code hosted from hulu.com.

Second, we found first party HTTP and HTML5 cookies respawned on hulu.com through a service hosted at kissmetrics.com. KissMetrics was exploiting the browser cache to store persistent identifiers via stored Javascript and cache ETags.

ETags are tokens presented by a user's browser to a remote webserver in order to determine whether a given resource (such as an image) has changed since the last time it was fetched. Rather than simply using it for version control, we found KissMetrics returning ETag values that reliably matched the unique values in their user cookies. To our knowledge, this is the first demonstration of this ETag tracking ''in the wild.''

ETag tracking and respawning is particularly problematic because the technique allows unique tracking even when the consumer blocks HTTP, Flash, and HTML5 cookies completely or enables private browsing mode. In order to block this tracking, the user would have to clear the cache between each website visit.

Conclusion: Why does this matter?

There are three privacy problems with the activities we detail in our report.

First, users cannot fairly be said to have notice of these activities. The entire point of new tracking methods seems to be motivated by users' ignorance of them. And the privacy policies we read didn't disclose Flash respawning or cache ETag use.

The lack of notice leads to a second problem: because these vectors are resistant to blocking, they rob con-

[10] Ryan Singel, *Online Tracking Firm Settles Suit Over Undeletable Cookies*, Wired Epicenter, Dec. 5, 2009, available at http://www.wired.com/epicenter/2010/12/zombie-cookie-settlement/.

[11] GUID typically means ''globally unique identifier.''

sumers of choice. This undermines the advertising industry's representations about respecting choice, and leaves consumers in a technical arms race with advertisers.

Third, there is much attention to whether websites *sell* information, but the first party tracking mechanism implemented by KissMetrics inverts the issue: how does tracking enable websites to *buy* information about their users from others?

The KissMetrics system uniquely enumerated users, and shared the same identifier with different first party sites (for instance, the same identifier beginning with ''GuTj890'' enumerated our browsing session at hulu, spotify, etsy, spokeo, and gigaom). This enabled these subscribers to KissMetrics to share information about users with other sites. Any of the above mentioned sites could ask each other for registration data about ''GuTj890.''

This is important because it breaks the trust model enabled by ''selective revelation.'' Advocates of market-based approaches to privacy have long argued that ''privacy is all about trust.'' Thus, the user ''trusts'' certain websites and shares only the amount of information that she is comfortable revealing in that context. For instance, a user may fear that Hulu.com would send spam, and thus provide a throw-away email address when signing up. That is a form of selective revelation that the market is supposed to respect.

However, if websites can simply go to information aggregators, selective revelation is no longer a workable strategy to protect privacy. Sharing any information—even fake information—could enable Hulu.com to match up cookies and discover real information that the user ''trusted'' to some other site. This risk is amplified where users are encouraged to authenticate in order to use a website's services, such as popular music or video services like Spotify or Hulu.

Finally, ''paternalism'' is an often-invoked canard against do-not-track and other privacy interventions. For instance, earlier this month, Thomas R. Julin argued in this publication (10 PVLR 1262, 9/5/11) that do-not-track interventions, ''. . . implement paternalistic judgments that subjects of targeted marketing cannot make proper judgments for themselves.''[12]

Participants in privacy debates will always be able to invoke ''paternalism'' as an objection to regulations.[13] But the real issue here is who gets to decide what the rules are. Governments may impose paternalistic privacy frameworks, but they recognize the dignity of individuals and attempt to balance their preferences against the needs of commerce. Advertisers see individuals as objects.[14] When conceived of as objects, consumers' preferences no longer matter. Privacy can be coded into oblivion or be circumvented with technology. Our 2009 and 2011 work empirically demonstrates that advertisers implement paternalistic judgments that subjects of targeted marketing cannot make proper judgments for themselves.

---

[12] Sorrell v. IMS Health May Doom Federal Do Not Track Acts, (10 PVLR 1262, 9/5/11) (Sept. 5, 2011).

[13] Hoofnagle, Chris Jay, *Denialists'* Deck of *Cards: An Illustrated Taxonomy* of *Rhetoric Used to Frustrate Consumer Protection Efforts* (February 9, 2007), available at http://ssrn.com/abstract=962462.

[14] Individuals are conceived of as ''targets'' and ''waste,'' Joseph Turow notes in THE DAILY YOU: HOW THE NEW ADVERTISING INDUSTRY IS DEFINING YOUR IDENTITY AND YOUR WORLD (Yale University Press 2011).

  PVLR ISSN 1538-3423