

Public Comment by Harlan Yu¹ and Ashkan Soltani²

On the Federal Trade Commission Staff Report: Protecting Consumer Privacy in an Era of Rapid Change

February 18, 2011

Consumers today do not have the means to adequately control their online privacy. Research has shown that during the course of ordinary Web browsing, consumers are unknowingly driven to a large number of third party online intermediaries that track their behavior across multiple websites.³ These tracking practices are widespread and often occur behind the scenes, without the knowledge of ordinary Internet users.

We believe that purely technical solutions are insufficient to protect consumers against these tracking practices and that our vision for Do Not Track will give consumers meaningful choice and control over these practices.

These comments, and our thoughts on Do Not Track as a whole, do not supersede calls for more baseline privacy protections. We believe that technical mechanisms and baseline policy protections need to work in tandem to improve online consumer privacy. In addition, by creating a mechanism that allows consumers to choose to stop online tracking, the Commission should also consider to extent to which online entities could compel consumers to reverse that choice and the implications of this possibility.

Do Not Track should focus on more than just online behavioral advertising.

For consumers, behavioral advertising is the most apparent use of the data collected about them through online tracking. However, there are many other potential uses of these data, unrelated to advertising, which may expose consumers to unexpected harm. Consumer dossiers are collected with few meaningful restrictions, and they can be sold or otherwise shared with downstream commercial entities, often without the consumer's knowledge or control. These dossiers may be used in contexts outside of online advertising and in ways that exceed consumer expectations of privacy for their online communications.⁴

¹ Harlan Yu is a Ph.D. candidate in Computer Science and the Center for Information Technology Policy at Princeton University.

² Ashkan Soltani is an independent researcher and consultant focused on privacy, security, and behavioral economics.

³ The KnowPrivacy Research Project. Joshua Gomez, Travis Pinnick, and Ashkan Soltani, "KnowPrivacy," June 1, 2009. <http://www.knowprivacy.org>.

⁴ The general public has relatively limited knowledge of the full extent that dossiers are shared and sold from one business to another. In one anecdote revealed by the Wall Street Journal, an insurance company was found to be making decisions using a "predictive modeling" system, based partly on consumer-marketing data..." See: <http://online.wsj.com/article/SB10001424052748704648604575620750998072986.html>, November 19, 2010. Given the obvious economic value of the data being collected, it

When consumer dossiers are collected and shared by a large number of online intermediaries and their affiliates, consumers are exposed to an increased risk of data breach. Major incidents of data compromise caused by accidental loss, malicious hacking, social engineering, or other reasons are well-documented⁵ and well-known to the FTC. Online tracking techniques today imply server-side data collection, and consumer risk increases as more copies of their data are made and distributed. Mere collection and retention poses risks to consumers, so consumers need better ways to communicate how they wish their data to be collected and used.

A Do Not Track choice mechanism based on an HTTP header.

The most elegant choice mechanism proposed thus far is based on a new HTTP header designed specifically for Do Not Track. The client sends the DNT header to simply inform the server about the user's tracking preference. The header is sent by the client with each Web request and can be set to two possible values: "1" if the user has chosen not to be tracked or "0" if the user has consented to tracking. If the header is not sent, the user has not expressed a tracking preference. Servers that receive the DNT header will need to honor the user's preference in accordance to FTC or industry self-regulatory guidelines.

To use the choice mechanism, consumers are presented with a simple and easy-to-locate checkbox—in the preferences or settings menu of the user client—to indicate whether they would like to be tracked online. On desktops, the user client is generally a Web browser. On mobile phones, the setting may exist at the mobile OS platform level such that the choice is reflected uniformly across all applications.

There are a number of advantages to the HTTP header mechanism.

First, the header mechanism will put an end to the technical arms race in online tracking.

Online entities today use a variety of methods to track users, and users are constantly scrambling to defend themselves against cutting-edge tracking techniques. Some methods are very difficult (if not practically impossible) for consumers to defend against, and new methods for tracking are continuously being invented.⁶ While there exists a patchwork of browser privacy tools, successful use of these tools requires advanced understanding about the privacy threat model and how the Web works. It also requires sustained effort to learn about newly invented tracking techniques, to find and install the latest tools, and to keep their tools up-to-date and well-configured. Needless to say, few

seems safe to assume that such unexpected uses of online tracking data already are or will be prevalent in the future.

⁵ See, for example, the DataLossDB research project (<http://datalossdb.org/>) and Verizon's 2010 Data Breach Investigations Report (http://www.verizonbusiness.com/resources/reports/rp_2010-data-breach-report_en_xg.pdf).

⁶ Standard Web cookies are just one method that servers use to track users and devices. There is a laundry list of other methods, such as flash cookies, browser fingerprinting (EFF's Panopticklick project: <http://panopticklick.eff.org/>) and the various methods used by Evercookie (<http://samy.pl/evercookie/>).

consumers make meaningful headway defending themselves against online tracking, and worse, successful use of the available tools may not fully protect even the most tech savvy consumers.

In contrast, under the DNT header framework, consumers need only notify servers of their preference not to be tracked, and compliant servers would comply with that choice regardless of which tracking method is used. This mechanism would relieve consumers from significant burdens that come from constantly trying to fend off the latest online tracking techniques. It also would allow companies to innovate with core business functions that maintain consumer privacy rights.

Second, the header mechanism is relatively easy to implement on both clients and servers.

It's been suggested that implementing Do Not Track in this way will require a substantial amount of additional work, possibly even rising to the level of "re-engineering the Internet." This is decidedly false. The HTTP standard is an extensible one, and it "allows an open-ended set of... headers" to be defined for it.⁷ Indeed, custom HTTP headers are used in many Web applications today.

On the client-side, adding the ability to send the DNT header is a relatively simple undertaking. For instance, it only took about 30 minutes of programming to add this functionality to a popular extension for the Firefox Web browser.⁸ Other plug-ins to send the header already exist.⁹ Implementing this functionality directly into the browser might take a little bit longer, but much of the work will be in designing a clear and easily understandable user interface for the option. Mozilla has already implemented the DNT header in their next major release of the Firefox browser.¹⁰

On the server-side, adding code to detect the header is also a reasonably easy task—it takes just a few lines of code in most popular Web frameworks.¹¹ It may take more substantial work to program how the server behaves when the header is "on," but some of this programming work has already been done. With industry self-regulation, compliant ad servers already handle the case where a user opts out of their behavioral advertising program—the difference now being that the opt-out signal comes from the DNT header rather than an opt-out cookie.

⁷ The IETF standard for the Hypertext Transfer Protocol -- HTTP/1.1.
<http://www.ietf.org/rfc/rfc2616.txt>.

⁸ Comment by Wladimir Palant on implementing the Do Not Track header for Adblock Plus, December 14, 2010.
<https://adblockplus.org/forum/viewtopic.php?f=4&t=6492#p40212>

⁹ The Universal Behavioral Advertising Opt-Out extension for Firefox by Christopher Soghoian and Sid Stamm (<https://addons.mozilla.org/af/firefox/addon/12765>) and the Do Not Track Firefox Jetpack Module by Jonathan Mayer (<https://github.com/jonathanmayer/Do-Not-Track/tree/master/jetpack/>).

¹⁰ Try Out the "Do Not Track" HTTP header. Sid Stamm, January 31, 2011.
<http://blog.sidstamm.com/2011/01/try-out-do-not-track-http-header.html>.

¹¹ Do Not Track: Web Application Templates. <http://donottrack.us/application.html>.

Note also that contrary to some suggestions, the header mechanism doesn't require consumers to identify who they are or otherwise authenticate to servers in order to gain tracking protection. Since the header is a simple on/off flag sent with every Web request, the server doesn't need to maintain any persistent state about users or their devices' opt-out preferences.

Third, the header mechanism maintains flexibility for existing commercial online practices.

When a consumer indicates that she wishes not to be tracked, servers will still be able to revert to practices that don't involve online tracking. For example, an advertising network can by default engage in targeted behavioral advertising, but for users who send the DNT header, the network can switch to contextual advertising instead. This header mechanism allows servers to determine the appropriate level of commercial service in each setting and continue to offer services to consumers.

Some proposed solutions to Do Not Track involve blocking HTTP connections to a list of purported tracking servers. Blocking solutions afford much less flexibility to online businesses, as they prevent them from providing any services at all to the consumer from listed servers. It's also not always easy for consumers to determine in advance which servers actually engage in tracking. In some cases, blocking lists will unknowingly "overblock" non-tracking servers and restrict otherwise legitimate business activity. In other cases, lists will "underblock" and fail to list certain tracking servers or be unable to keep up with the continuous introduction of new tracking servers.¹²

Designing a clear and usable interface for the mechanism.

Do Not Track is a powerful idea because it is simple. It needs to be usable by all Internet consumers regardless of their technical sophistication. The primary interface should be a prominent, logically-placed checkbox option in the user's client. When the option is turned on, the client should initially send the "enabled" DNT header to every site.

It will likely be necessary to provide advanced users with more granular controls. For example, a consumer may want to later consent to tracking on a limited basis, while declining tracking to all other parties. These controls and indicators could be built into the advanced preferences menus or within the context of the content body.

With flexibility comes complexity, and interface designers will need to be careful about not making the mechanism too complex for ordinary consumers to use. Client vendors and privacy researchers should conduct usability studies to determine how the mechanism's user interface should best be designed.

¹² It would not be difficult for tracking entities—if they desire—to "engineer around" blacklist solutions. In a useful historical corollary, DNS blacklists (<http://www.dnsbl.org>) were created and used by many email spam filtering platforms to block unsolicited emails. While initially successful, many spammers were able to circumvent these technical limitations, by constantly migrating domains from which they sent spam, along with other techniques.

Defining what “tracking” means.

Whether or not a universal Do Not Track mechanism is eventually developed, the industry and the FTC should decide on a meaningful definition of what protections consumers receive when they choose to not be “tracked” online. As stated above, online tracking poses harms to consumers outside the realm of behavioral advertising. The National Advertising Industry’s self-regulatory opt-out mechanism only promises that their members will “no longer deliver ads tailored to your Web preferences and usage patterns.”¹³ This is insufficient to protect consumers from non-advertising-related privacy harms posed by increased data collection and retention.

The definition of tracking should be based not just on data use, but also on collection and retention. In particular, the definition should focus on the practice of building consumer dossiers or profiles, whether or not the entity is able to attach a consumer’s real-world identity to the profile. It should also be based on reasonable consumer expectations of privacy.

In nearly all first party settings, consumers should reasonably expect that the entity in the browser’s URL location bar may retain data about all interactions on that site. Consumers should also expect that third party entities they interact with directly—for instance, when clicking on an advertisement or interacting with an embedded social widget—would track the interaction. However, consumers would likely not expect tracking by the mere presence of a third party widget (absent direct interaction) even if that widget is tied to an authenticated session.

While it may be clear to some consumers that ads or other separately branded elements are third party in nature, many loaded objects on a website, such as social widgets or embedded videos, may mistakenly be perceived as first party interactions. The definition of third party tracking should attempt to address some of these nuances in order to avoid accidental bias toward certain market players.

The first and third party delineation of online entities should be based on more than just the strict technical separation based on the domain entity that appears in the location bar. Consumer expectations of data collection and sharing are also based on known business relationships and branding, so such separation should attempt to follow these same general contours.

There should be a clearly defined set of “commonly accepted commercial practices” that are explicitly exempt from the definition of tracking. In particular, many common logging practices—for example, to defend against security breaches or click fraud—should be allowed, as long as reasonable data retention policies are in place and the logged data are not deliberately associated with any commercial consumer profiles.

¹³ The NAI’s Behavioral Advertising Opt Out Tool:
http://www.networkadvertising.org/managing/opt_out.asp

Enforcing and detecting violations.

One initial way to increase confidence in the DNT mechanism is for a server to acknowledge that it received the DNT header in its HTTP response. This will help consumers verify that the header was actually received by the server and not accidentally stripped by network intermediaries (for example, a proxy server). It may be possible to add additional signaling mechanisms between the client and the server in the standards for the DNT header protocol.

Recently, there has been a move by some industry groups to provide ‘in ad notifications’ that allow consumers to better understand how each individual ad was chosen and opt-out of specific ad companies or categories.¹⁴ One great benefit of these ‘in ad notifications’ is that they contain or link to metadata describing the underlying mechanisms of how a given ad was served (e.g. based on personalized behavioral data or contextually based on the content of the site). This metadata could potentially also be used to detect when specific companies are not respecting the DNT header, either maliciously or through system error. Note that this doesn’t address bad actors or those that do not participate in these ‘in ad notifications’. However, it does provide a way to “debug” the ones that do.

For those advertisers and trackers that fall outside of these voluntary methods, there are data-driven approaches that can be employed to detect some types of tracking. For example, sampling ad distribution for users with and without a DNT header may allow a compliance entity to compare differences and infer, with some level of confidence, whether or not ads were shown based on random, contextual, or behavioral data. There are a number of challenges to measuring online ad systems in this way and academic research in this area is still young.¹⁵

Ultimately, effective enforcement may require individual audits of companies that are believed to not be in compliance with the DNT header. Web server logs may show the presence of a behaviorally targeted advertisement that was served despite a DNT-enabled request. A source code audit may also reveal offending behaviors. In the limit, invisible trackers (e.g. web bugs) may not exhibit any user-facing behavior and may require these kinds of internal investigations of the entity’s back-end data collection practices.

One inherent weakness in any FTC regulation is that the FTC only has jurisdiction over entities based in the United States. While this is the case, there is reason to believe that FTC regulation would have significant influence over online tracking practices as a whole. Surveying the most pervasive online tracking entities as of March 2009, we found that 88 of the top 100 trackers—including 29 of the top 30—are based in the United

¹⁴ Evidon InForm (http://www.evidon.com/assurance_platform) and the IAB/NAI CLEAR Ad Notice Technical Specifications (http://www.networkadvertising.org/pdfs/CLEAR_Ad_NoticeApril2010.pdf).

¹⁵ Saikat Guha, Bin Cheng and Paul Francis. Challenges in Measuring Online Advertising Systems, in the proceedings of the 2010 ACM Internet Measurement Conference. <http://saikat.guha.cc/pub/imc10-ads.pdf>.

States.¹⁶

Beyond the browser.

Much of the tracking debate extends to areas beyond just a standard Web browser used on a consumer PC. As devices, such as mobile phones, set-top boxes, and even automobiles become “Internet enabled,” it may be possible to extend the Do Not Track mechanism to these devices. Mobile online platforms are in the nascent stages of development and often lack even basic privacy mechanisms that standard browsers provide, such as the ability to delete cookies or use privacy-enhancing plug-ins. Additionally, mobile platforms and embedded devices provide advertisers with even more persistent ways to track users since they often transmit hardware device identifiers to advertisers.¹⁷

Appendix A.

The list of the most prevalent trackers found on the 100 most popular websites used here was gathered by the KnowPrivacy research project in March 2009. We looked up the corporate location for each tracking entity using a variety of sources, including Ghostery company profiles, the WHOIS database and information on the entity’s corporate website. We show the list of trackers, in order of prevalence, with its country origin below:

¹⁶ For the detailed list of trackers and corporate locations, see Appendix A.

¹⁷ Your Apps Are Watching You. The Wall Street Journal, November 19, 2010.
<http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html>.

1. USA -- Google Analytics
2. USA -- Doubleclick
3. USA -- Microsoft Atlas
4. USA -- Omniture
5. USA -- Quantcast
6. USA -- PointRoll
7. USA -- Google AdSense
8. USA -- Dynamic Logic
9. USA -- InsightExpress
10. USA -- ValueClick Mediaplex
11. USA -- AddThis
12. USA -- Revenue Science
13. USA -- RightMedia
14. USA -- Zedo
15. USA -- SpecificClick
16. USA -- Tacoda
17. USA -- WebTrends
18. USA -- DiggThis
19. Canada -- Casale Media
20. USA -- NetRatings SiteCensus
21. USA -- Tribal Fusion
22. USA -- Google Custom Search Engine
23. USA -- Quigo AdSonar
24. USA -- Rubicon
25. USA -- Yahoo Buzz
26. USA -- Facebook Connect
27. USA -- AdBrite
28. USA -- Burst Media
29. USA -- Twitter Badge
30. USA -- MSN Ads
31. USA -- OpenAds
32. USA -- SiteMeter
33. Ireland -- Statcounter
34. United Kingdom -- Vibrant Ads
35. USA -- Wordpress Stats
36. USA -- Chitika
37. USA -- Technorati Widget
38. USA -- Google Widgets
39. USA -- ShareThis
40. USA -- Sphere
41. USA -- FeedBurner
42. USA -- MyBlogLog
43. USA -- Adify
44. USA -- CPX Interactive
45. USA -- Crazy Egg
46. USA -- Feedjit
47. USA -- Snap
48. USA -- Amazon Associates
49. United Kingdom -- Clicky
50. USA -- Lotame
51. USA -- AddtoAny
52. USA -- Baynote Observer
53. USA -- BlogCatalog
54. Hungary -- Blogads
55. USA -- Digg Widget
56. USA -- Disqus
57. USA -- Google FriendConnect
58. USA -- JS-Kit
59. USA -- Kanoodle
60. USA -- Kontera ContentLink
61. USA -- Lijit
62. USA -- AdaptiveBlue SmartLinks
63. USA -- Advertising.com
64. USA -- Alexa Traffic Rank
65. Canada -- Clicksor
66. USA -- Federated Media
67. USA -- HitTail
68. USA -- LivePerson
69. USA -- ScribeFire QuickAds
70. USA -- TriggIt
71. Netherlands -- TwitterCounter
72. USA -- Yahoo Overture
73. USA -- YieldBuild
74. Brazil -- BTBuckets
75. Germany -- BlogCounter
76. USA -- BlogHer Ads
77. USA -- ChartBeat
78. Israel -- ClickTale
79. USA -- Criteo
80. USA -- Crowd Science
81. USA -- Cubics
82. USA -- FriendFeed
83. USA -- HitsLink
84. USA -- IndexTools
85. USA -- Intense Debate
86. USA -- Lookery
87. USA -- Omniture TouchClarity
88. USA -- Others Online
89. USA -- Outbrain
90. France -- Piwik Analytics
91. Canada -- PostRank
92. USA -- Reinvigorate
93. USA -- Salesforce
94. USA -- Six Apart Advertising
95. USA -- Statify
96. USA -- UserVoice
97. USA -- W3Counter
98. USA -- WidgetBucks
99. USA -- Woopra
100. USA -- AdultAdWorld

