

**From:** "Rosenthal, Noga" <Noga.Rosenthal@247realmedia.com>  
**Subject:** **RE: Expression of Interest: W3C Workshop on Web Tracking and User Privacy**  
**Date:** 18 April 2011 20:30:12 GMT+02:00  
**To:** Thomas Roessler <tlr@w3.org>  
**Cc:** "team-privacyws-submit@w3.org" <team-privacyws-submit@w3.org>  
▶ 2 Attachments, 2.6 KB

---

To the World Wide Web Consortium (W3C):

24/7 Real Media, Inc. provides digital marketing solutions for advertisers and publishers through its award-winning ad management platform, Open AdStream®, and the Global Web Alliance, which consists of high quality publisher websites.

24/7 Real Media, Inc. is firmly committed to protecting the privacy of Internet users and fostering user confidence in the Internet. We are dedicated to following the [Code of Conduct](#) of the [Network Advertising Initiative \(NAI\)](#), of which we are a member. We are also members of other self-regulatory organizations, including the Interactive Advertising Bureau, both in the US and the UK.

24/7 Real Media's commitment to user privacy make us a material stakeholder in the discussion regarding the delivery of online behavioral advertising and user privacy. Therefore, we would like to participate in the W3C Workshop on Web Tracking and User Privacy.

Regards,  
Noga Rosenthal



**Noga Rosenthal** > Corporate Counsel  
T: +1.212.231.7265; F: +1.212.760.2811  
noga.rosenthal@247realmedia.com

24/7 Real Media, Inc., A WPP Company  
132 West 31st Street - 9th Floor  
New York, NY 10001

➔ The Science of Digital Marketing  
[www.247realmedia.com](http://www.247realmedia.com)

# Objectives for W3C Work on Web Tracking and User Privacy

W3C Workshop on Web Tracking and User Privacy

28/29 April 2011, Princeton, NJ, USA

Submitted by: AT&T

Contributors: Bryan Sullivan  
Ileana Leuca  
Sherry L Ramsey  
Michael Merritt

---

## 1. Introduction

Past efforts at privacy-enabling standards (e.g. P3P) have reinforced that specifications alone don't solve problems. W3C should thus first facilitate a dialog on Web tracking and user privacy which establishes clear objectives addressing the Web as an ecosystem, part of an overall services marketplace in which Web tracking has a role, and in which user privacy is a clearly defined and achievable goal. It may take some time for an iterative process of specification, prototyping, and deployment experience to achieve a workable balance between the needs for tracking (e.g. for marketing and service personalization) and the desire for privacy. But that timeline will be shortest only if a comprehensive, shared understanding is first achieved on:

- The roles and objectives of tracking and marketplace stakeholders
- Characteristics of a desirable solution balancing tracking and user privacy
- The limits of current technology to achieve the desired solution

## 2. Roles and objectives of tracking and marketplace stakeholders

The Web is an ecosystem within a larger services marketplace, in which marketing data collection and service personalization are examples of how user/service information is used. Before considering detailed requirements or technologies supporting possible solutions, W3C should first:

- Seek consensus on the actual (i.e. current) role and methods of tracking in the Web ecosystem: Tracking does have an actual role, whether one considers it necessary or undesirable, and that role needs to be understood prior to

implementation of privacy-enabling solutions, to prevent undue negative effects on the Web ecosystem. Further, the methods currently used for tracking will need to be understood to ensure that privacy-enabling solutions are effective.

- Frame the role of tracking and desire for privacy within the set of roles and objectives of marketplace stakeholders, including:
  - Users, including individuals, families, and enterprises
  - Web user-agent developers, for browsers and other Web-enabling runtime environments (e.g. W3C widget runtimes)
  - Application developers, for client and server based applications
  - Service providers, including network service providers and Web service providers

If there is a role for “privacy/trust certification” as part of what W3C recommends, this should support market-based/globally-applicable approaches, e.g. as with existing PKI services and trusted application distributor models, e.g. for which privacy certification can be an aspect of overall user safety based upon trust in the application distributor.

### 3. Characteristics of a desirable solution balancing tracking and user privacy

The most important lesson learned from earlier efforts at W3C privacy standards is that W3C should standardize what has been proven to work in the market, i.e. has been developed, deployed, and used for some time, successfully. Standardization of technologies should not occur first – rather, objectives and guiding characteristics for solutions (including technology choices, where necessary) should be established, and quickly prototyped by Web user agent developers. This will necessarily require an iterative process of specification, prototype, and deployment experience, before a final technology standard is achieved.

Some objectives and guiding characteristics for solutions should include support for:

- an overall good user experience, e.g. easy to use, whether one wants to “opt-in” or “opt-out” by default, and change preferences easily and quickly as conditions warrant
- context adaptability, e.g. works well in different types of devices and user-agents, and whether a user’s own or borrowed device is used
- effectiveness, e.g. resulting in a real sense of enhanced privacy
- limited impact on the Web services marketplace, e.g. does not “break the Web” or overly impact existing Web business models
- a technology basis in the W3C’s existing content formats and user-agent behavior specifications, e.g.
  - HTML5, CSS, DOM
  - [POWDER](#), e.g. as extended by [WAC](#) in “[WAC 2.0](#)”, to address “[Privacy Considerations for API Usage](#)” and “[Privacy Considerations for Device Property Access](#)”

## 4. The limits of current technology to achieve the desired solution

It is important to understand the limits of technology to address the overall objectives of user privacy, in order to set reasonable expectations on what types of protection can be provided. For example, goals for user privacy may include that the user is always able to:

- Know that use of their information is actually limited to the disclosed use
- Know what information has been shared with whom (including who they have shared it with, etc), and where it still exists
- Revoke access to private information (including that which has already been shared, collected, or stored) and capability to request removal of retained individually identifiable information or be assured that it has been anonymized or aggregated.

These goals however are only partially achievable with current technology. At most it may be possible to express intent for private data use/exposure, and consent of the user to that intent. Verification of actual compliance to the stated intent and consent may require unspecified audit processes.

## 5. References

[POWDER](#): Protocol for Web Description Resources (POWDER), W3C

[WAC 2.0](#): WAC 2.0 Proposed Release Version (PRV), Wholesale Applications Community (WAC)



## **Do Not Track: An Outcomes Analysis**

Author: Andrew Sudbury ([andrew@getabine.com](mailto:andrew@getabine.com)) and Rob Shavell ([rob@getabine.com](mailto:rob@getabine.com)) co-founders, Abine, Inc.

### **Abstract**

First, the bar the industry collectively sets for the percentage of consumers who understand how they're tracked online can and needs to be far higher. Second, because today's operators have free reign to track consumers clandestinely by default, the advertising and publishing industry should bear costs to significantly raise the bar – the percentage of consumers who understand online tracking (for example, the percentage who can answer simple questions about how they're being tracked). Third, this goal can be accomplished quickly by introducing incentives to pair advertising / publishing industry players and consumer protection / privacy providers together to jointly develop and test solutions that measurably increase transparency and awareness.

### **Abine's Position on Do Not Track**

Abine, Inc. The Online Privacy Company, is a leading provider of online privacy solutions for consumers. Among our offerings actively being used in the market today is a suite of consumer tools designed to mitigate online tracking by allowing users to opt out of ad networks, delete cookies, enable IP masking proxies, create single-use phone numbers and email addresses, and block Javascript, pixel images and referrers.

Technology that tracks online activity has become more advanced and invasive, and, in response, Abine has had to keep pace through constant readjustment of our tactics. This experience of advocating for online privacy in a constantly shifting landscape has made us wary of simple one-size-fits-all solutions. Rushing to implement "Do Not Track" via technology or policy or both could result in knee-jerk behavior where consumers hear an appealing phrase they don't adequately understand (and equate to not getting telephone calls) and to then blindly insist they never be tracked online.

We believe it is important for *all* parties involved in online tracking to focus on outcomes, rather than on specific technologies and that the first outcomes data should illustrate / increase the level of understanding consumers online have of how they're being tracked. Increasing understanding first, and then measuring metrics generated by consumer solutions which balance privacy and personalization differently, is a good way for a consumer-friendly yet reasonable set of choices to emerge.

Our strong view is that firstly, the bar for consumer understanding of online tracking needs to be far higher. Secondly, because today's operators have free reign to track consumers clandestinely by default (a.k.a. opt-out) that the advertising and publishing industry should bear the costs of significantly raising the percentage of consumer understanding of online tracking (for example, the percentage who can answer simple

questions about how they're being tracked). Thirdly, that goal can be accomplished quickly by introducing incentives to pair advertising and publishing industry players and consumer protection / privacy providers together to jointly develop and test solutions that increase awareness levels.

### **Do consumers understand Tracking and Do Not Track?**

No. As we put this paper together, we did a flash survey of 250 respondents with a sample set of consumer tracking awareness questions. Here are the results:

In response to: "Are you tracked when you surf and visit web sites?"

- 47%** (82 Votes) Sites I visit know my internet connection (IP address) and track that.
- 20%** (35 Votes) Sites I visit track me and send my information to a network of other companies and advertisers.
- 18%** (31 Votes) Sites I visit know my specific computer and have information about other sites I visited also.
- 14%** (25 Votes) Sites I visit can only track me until I turn off my computer.

However anecdotal, this data is in-line with our experience rolling out and supporting different interfaces designed to make Web tracking more transparent to hundreds of thousands of online users. Consumers are unaware of the extent of the different ways they're being tracked.

The simple goal should be for a high degree of awareness, combined with a low amount of frustration / friction in return for the knowledge. Downstream, additional outcomes to measure would include those that show users are making more informed and nuanced choices than "on or off" (for example, evidence showing they are making their own decisions more frequently on a case-by-case basis) or those that show users are trusting third parties to do so on their behalf (think Microsoft's TPL's or in domain further afield, the Lifelock service).

While it's nothing new to suggest users need to understand information in order to make informed choices, we believe some much higher percentage than today (10%? 25%?, 50%,?) should have a base level of awareness, as determined by qualitative surveys. A lot of this responsibility should rest with industry operators.

### **Industry operators should pay (some) costs related to Do Not Track**

It seems premature for lawmakers or consumer advocates to demand the advertising / publishing industry bear unknown levels of compliance costs for a deceptively simple Do Not Track header / preference expression by implementing different technology workarounds across thousands of already-implemented and operational systems. This is leaving aside economic arguments concerning loss of value from targeting which are empirically unconvincing, but out of scope.

Industry should have to bear the costs of communicating tracking transparently to users of its products and services, no matter where a publisher, ad network, or data analyzer sits

in increasingly complex value chain. We see far too many arguments from insiders that pass the buck to others and rely on IP address logs = anonymity. These are mostly invalid and trite to say the least. The industry owes to the consumer this significant—but obtainable—level of transparency to facilitate awareness in exchange for being able to operate by default as opt-out, e.g. in its current unobtrusive manner.

Furthermore, industry should not be able to define what transparency means (e.g. a logo on advertisements) just like cigarette manufacturers don't decide the size and language of the warning labels they must display. An operational definition of transparency is most likely to be devised in one of two ways: either an impartial party should decide what is a reasonable standard or parties with competing agendas with an incentive to collaborate.

We observe there to be scant evidence that self-regulatory approaches deliver meaningful results. Neither a privacy policy's legalese, nor certifications by third parties with trust seals, nor in-ad icons proposed by the self-regulated, would meet our proposed simple survey-based results threshold for effective communication – whether that is: 10%, 25%, or 50% of consumers being able to state with any accuracy how they are being tracked and if they are ok with that, e.g. “tracking awareness”.

We believe transparency should be measured by overall user awareness metrics and that improving tracking awareness may come down to providing promotional “shelf space,” which may represent a price too high for browser vendors, publishers, and advertisers to pay without additional motivating factors. For example, Microsoft, perhaps the best of the browser privacy promoters today, announced Tracking Protection Lists for IE9 (TPL) a privacy feature nested inside a menu choice called “Safe” – Microsoft could have provided a way for users to see the available list of vendors providing IE9 tracking protection choices. They did not.

### **Getting consumer-friendly and measurable solutions to market**

Once a basic level of consumer awareness is reached, any subsequent choice should be respected and enforced (and tracking customers with the intent of respecting these choices about further tracking should be encouraged). Consumer exposure, choice, and the persistence of that choice across sites and advertisers is likely the most complex terrain for participants to navigate. Dangers include overly simplistic on/off settings, obtrusive pay-walls, and fatiguing over-communication.

Though it's possible to create a system that's ultimately unnecessary in the marketplace, the value of future privacy solutions can be ensured by paying close attention to a basic set of outcomes metrics and by creating incentives for the continued development of solutions through a collaboration of the different players in the privacy marketplace, many of whom would not naturally partner to create these kinds of experiments without solid economic reasons to do so.

We'd suggest that the advertising industry and publishers pay the price of advertising solution alternatives to tracking to their users, and that these alternatives be provided by an innovative marketplace of companies and organizations, rather than by advertisers and first-party publishers alone or by a regulatory body

Aware consumers should bear costs as well, of course. It's reasonable to ask those consumers who wish not to be tracked to pay for privacy-by-default, rather than push those costs to industry by default (Do Not Track). After all, by default, consumers are accessing mostly free services they find valuable. Furthermore, if industry violates the wishes of a consumer paying for increased online privacy, there is then monetary harm which can help to establish concrete damages.

## **Summary**

Instead of threatening regulation or infighting unnecessarily, the parties involved in reforming online tracking should work together to enable a healthy market system that is responsive to consumer outcome metrics. It seems ironic in an industry full of measurement and tracking, the discussion on curtailing these includes so little relevant data. Such data could be obtained and improved easily today, especially by pairing currently-available solutions in the privacy market with existing publishers and advertisers. With the right focus, a set of viable Do Not Track metrics based on responses to existing technologies via experiments with distribution partners could make measurable strides which benefit consumers everywhere.

# Position paper for the W3C Do Not Track Workshop

## Aleecia M. McDonald

---

### Summary

Preliminary research suggests that user's expectations for Do Not Track (DNT) will not match implementations. While we might imagine changing DNT implementations to align more closely with expectations, it is quite unlikely DNT will change enough to meet user expectations. For example, if users think Do Not Track means no data collection at all, advertisers are unlikely to forgo counting unique clickthrough rates for billing. Furthermore, it is likely there will be multiple approaches to what Do Not Track means in practice, creating additional user confusion and uncertainty. Communicating with users to explain the gap between their expectations and reality is crucial. That means creating mechanisms to support (or at least not preclude) DNT implementers explaining how they implement DNT, and what their implementation means to their users. Unfortunately, current standards proposals do not envision this type of feedback to users. I hope to spark discussion about expanding DNT standards to include the data required to communicate with users.

### Established Issues

Users do not understand the Network Advertising Initiative (NAI) description of their members' opt-out cookies. In research from Carnegie Mellon's CUPS laboratory, we presented a screenshot of the NAI website and found only 11% of study respondents selected the correct multiple-choice description of NAI opt-out cookies. Our largest group of respondents mistakenly believed their data would not be collected if they opted out.<sup>1</sup>

Part of the confusion with NAI opt-outs may stem from the multiple ways in which NAI members implement opt-outs. Some OBA companies stop collecting data when they read opt-out cookies. Some companies, including Google, aggregate data from all users who opt-out. Some companies, including Yahoo!, do not change their data collection practices. They stop showing ads tailored based on user data, but data collection continues unchanged. So much variation in outcomes poses a difficult communication problem. There is no one, simple answer to the basic question: what does an opt-out cookie do?

---

<sup>1</sup> McDonald, A. M., and Cranor, L. F. Beliefs and behaviors: Internet users' understanding of behavioral advertising. In *38th Research Conference on Communication, Information and Internet Policy* (Telecommunications Policy Research Conference) (October 2 2010).

In addition, when users see a checkbox labeled “opt out” next an advertiser’s name, they are likely to expect they are opting out of seeing advertising from that advertiser. NAI takes great pains to stress that users will see the same number of ads with or without opting out, perhaps because NAI had discovered this is a common misconception. But even with a warning in bold that opt-outs do not reduce ads, we still found that was a common misconception. It is even more difficult to communicate clearly when users hold an expectation that does not match the implementation for privacy controls.

There are three ways in which NAI opt-out cookies research is directly relevant to Do Not Track. First, Google’s Chrome browser uses opt-out cookies as their Do Not Track solution. Presumably they have similar communication challenges with their users as the NAI has had. Second, the Do Not Track header sent by both Firefox and Internet Explorer will likely to encompass multiple implementations, as different parties define “tracking” in different ways. Most immediately, some companies may initially treat the DNT header exactly as they do opt-out cookies, thus recreating all of the ambiguity already inherent in opt-out cookies. Third, preliminary research strongly suggests when users see the phrase “Do Not Track,” they mistakenly believe this means all data collection stops. As with opt-out cookies, when users think they understand what something means, but it turns out to mean something else, there is a challenge to communicate across the gap between user expectations and reality.

## Mind the Gap

As just one example of how complicated defining “tracking” has become, Figure 1 contains a list of data uses that the Center for Democracy and Technology consider to be tracking, or not.<sup>2</sup> To understand this chart, users would need to understand at least the difference between first- and third-party websites, what behavioral advertising is, the types of data collected for behavioral advertising, the difference between identifiable and non-identifiable data, reporting, and analytics.

In a pilot test for a larger on-going research study, I found a majority of users expect Do Not Track to eliminate all data collection. The study starts by asking participants what they expect a Do Not Track button in their web browser would do. Participants work from their own expectations rather than a definition of DNT. They check the types of data they believe can be collected before and after clicking a Do Not Track button, with a subset of results shown in Figure 2.

---

<sup>2</sup> Center for Democracy & Technology. What does “Do Not Track” mean? A scoping proposal by the Center for Democracy & Technology, January 2011. <http://cdt.org/files/pdfs/CDT-DNT-Report.pdf>.

<b>Tracking</b>	<b>Not tracking</b>
Third-party online behavioral advertising	Third-party ad and content delivery
Third-party behavioral data collection for first party uses	Third-party reporting
Third-party behavioral data collection for other uses	Third-party analytics
Behavioral data collected by first parties and transferred to third parties in identifiable form	Third-party contextual advertising
	First-party data collection and use
	Federated identity transaction data
	Data collection required by law and for legitimate fraud prevention purposes

Figure 1: The Center for Democracy & Technology’s list of examples of data used for tracking and not tracking, illustrating their definition of tracking

To highlight a few of the more interesting results in the pilot study:

- 61% of respondents expected that if they clicked a Do Not Track button, websites would collect no data at all. None of the current proposals for Do Not Track contemplate limiting data collection to nothing for first party use, yet that is what many users expect from Do Not Track.
- Respondents did not expect Do Not Track to work by aggregating their data with other user’s data, with only 5% selecting that as a possibility, yet this is how some companies treat opt-out cookies today. Similarly, participants did not expect Do Not Track to work by collecting the same information, but anonymizing it, with only 7% selecting that as a possibility. One reason participants may not expect DNT to protect privacy via aggregation is because they believe that is already how the Internet currently works, and do not understand that they are uniquely identified today.
- Only 7% of respondents expected that websites could collect the same data before and after users click Do Not Track. Some DNT implementations may limit data use rather than data collection, as Yahoo! does with opt-out cookies today.

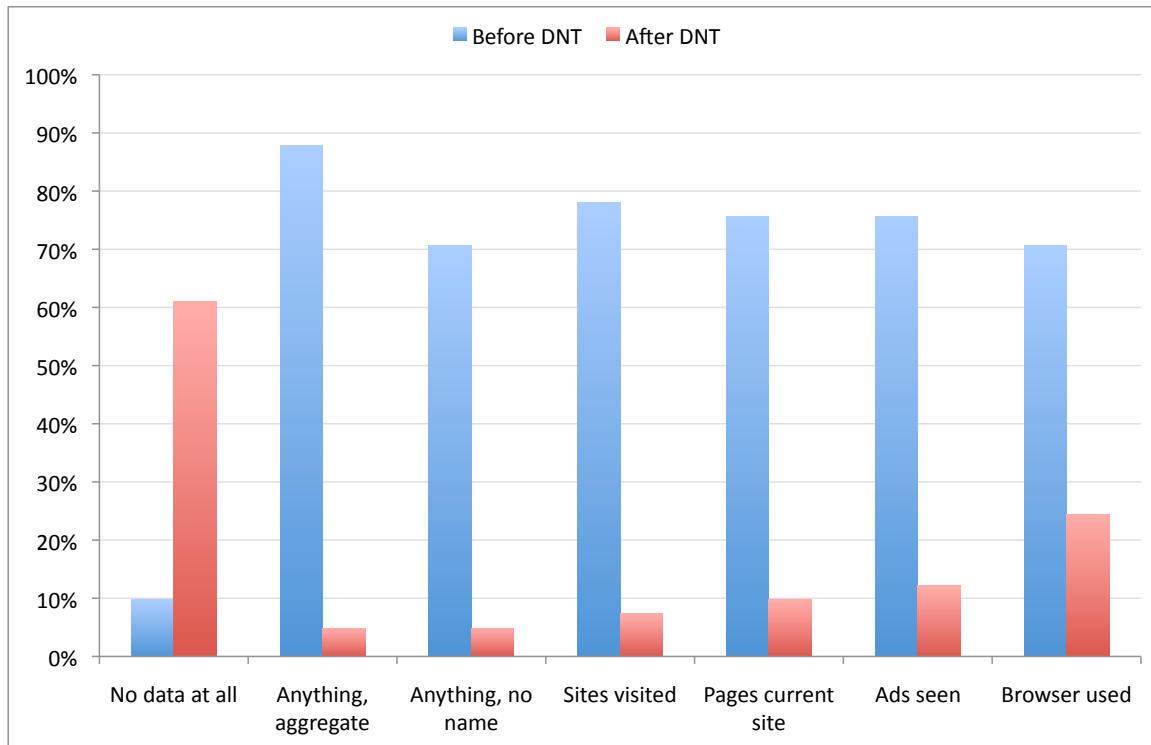


Figure 2: Types of data users think websites can collect, before (in blue) and after (in red) clicking a "Do Not Track" button in their web browser

A larger study is currently underway, and results will be ready for discussion at the W3C workshop. The high level conclusion should remain stable: when users hear "Do Not Track," the majority of users believe data collection stops.

## The Role for Standards

One way to address the gap between user expectations and reality is to communicate what DNT actually does. Existing mechanisms include online help files from browser makers and privacy policies from DNT implementers. However, most users do not read online help or privacy policies, and do not understand the number of entities collecting data about them on any given website.

The current DNT proposal before the IETF contemplates browsers sending a DNT header, and receiving confirmation of what the browser sent.<sup>3</sup> Beyond simple acknowledgment of the DNT header, this standard omits any automated mechanism for DNT implementers to communicate with end users. If instead standards build in communication channels, we can create transparency around what DNT means for any given DNT implementer. With multiple DNT implementers rolled up together, we can provide a holistic view of privacy implications for a particular visit to a

<sup>3</sup> J. Mayer, A. Narayanan, S. Stamm. Do Not Track: A Universal Third-Party Web Tracking Opt Out, IETF Draft (March, 2011).



particular website. Finally, we can create an opportunity for sites to communicate the benefits of personalization and why they use data.

If the W3C were able to agree upon standards for communication about DNT implementation details, IETF might extend the current proposal. Or, there may be better forms of communication that do not require modifying HTTP headers. As an example, an extended standard DNT response might include:

- An acknowledgement of the DNT header, as currently proposed to the IETF
- A URL with human-readable text describing what Do Not Track means to that particular entity. This could be as simple as an anchored tag in a privacy policy, for example [www.acme.com/privacy.html#dnt](http://www.acme.com/privacy.html#dnt)
- A standardized code describing DNT practices, as below

While privacy policies have too much variation to fit neatly into a handful of pre-defined categories, DNT implementations may be more tractable. For example, a site might be classified as type-0 if it only implements DNT by suppressing targeted ads but otherwise continues data collection and use identically, type-1 if data is aggregated, and type-2 if data collection stops all together. If that proves too simplistic, a code could be created from binary values for an ordered set of categories, for example 101 might mean a site collects data for fraud prevention, does not show targeted ads, and does collect data for analytics. It is not necessary to define a standardized code at this early juncture. However, it would be exceedingly helpful to think about what the syntax might look like, and build in a mechanism to support communication.

Eventually, user agents could use this data to inform users about their effective privacy online. However, if we create standards that preclude information flow, we will not be able to build visualization tools later.

I would like to speak at the W3C workshop to highlight the gap between user expectations and what is being built, explain that feedback to users can help communicate across this gap, and encourage discussion about how best to include feedback mechanisms in DNT standards.

# *Privacy and the W3C: principles and questions*

For the W3C Privacy Workshop, Princeton, April 2011

**David Singer**

*Multimedia and Software Standards, Apple Inc.*

## **1 Overview**

“What privacy specification work should the W3C perform?”

This paper first looks at the question of why the W3C should be active. It then asks some questions about the general principles involved, and finishes by examining three areas: privacy policies, the privacy aspects of other specifications, and the recent W3C member submission [1], including “Do Not Track” [2].

## **2 Why the W3C should be involved.**

Privacy often becomes implicated when either state is involved (the recording of user information, in particular), or the integration or correlation of services, or both. The W3C is the owner both of specifications that handle state, and is the owner of the integrated ‘web platform’. It is uniquely placed to handle the privacy implications in these areas.

## **3 Principles**

Almost everything that affects this area is in flux:

- user understanding, expectations, and perceptions;
- the ‘web platform’ – the protocols and formats used;
- business models, and business activity and services.
- legislative and regulatory activity, and social norms;

We somehow have to expect, and allow for, ingenuity and invention in business and services, and evolving user understanding and possibly even expectation, yet also develop specifications and policies that attempt ‘minimal surprise’ to users.

Users seem often upset by surprise, reacting negatively when something happens – even if they would have consented if asked in advance. Surprise is sometimes compounded when policies and similar documents are ‘long’ or ‘complex’, and users consequently do not fully understand (or perhaps even read) them.

Both the web platform and business practices and techniques are evolving; purely technical statements (e.g. “do not use HTTP cookies”) are liable to fail to manage new techniques, while purely effect-based statements (e.g. “do not ‘tag’ or track the user”) are

liable to interpretation and hence disagreements over interpretation. We will probably need a balance of the two.

The techniques needed to handle the truly intrusive sites are much heavier than most users would want to use most of the time. The W3C usually assumes a co-operating community, and it may be best to focus on that area initially, and leave ‘protection against the hostile’ to developers, and the future.

Finally, there is a balance needed between positive and negative effects. For example, users may not use a technique if the most visible effect is that services immediately stop working (even if there is a less-perceptible long-term benefit). Web services may not use a technique if they perceive the most likely short-term outcomes would be negative or neutral, even if there is a long-term benefit. It is critical that users see a benefit both to using and not using this request, and that services see a benefit to themselves as well as to the users, in responding.

## **4 Privacy Policies**

Making privacy policies short and clear helps reduce the possibility of misunderstanding, and helps a goal of ‘minimal surprise’. Two actions might help promote shorter and/or better understood policies.

The first is establishing definitions of terms. The ITU has a specification [4] that defines some of the terms used in this area; however, web-specific terms are not included (e.g. “cookie” is missing). The IETF also has a document [5]. The W3C might usefully publish definitions and ‘background information’ on web-specific terms.

The second is establishing a ‘database’ of common policy fragments. For example, the W3C might identify a few ‘legal disclosure’ policies, and give them names. This would enable corporate policies to say simply “our legal disclosure policy is W3C-Strict [ref]”. This, in turn, permits users to make decisions, or requests of their user agents, or enables user-agents to provide succinct summaries (e.g. using privacy icons [8]).

## **5 Existing Specifications**

The IETF has a draft under way that explores the privacy considerations of implementing Internet protocols [3]. There are also privacy implications of implementing W3C specifications. At the moment, we are not doing a systematic review of specifications to explore their privacy implications. In the past, this has led to problems such as the famous CSS link-visited issue [9]. HTML5 has a number of new state-handling techniques, which clearly have privacy implications (such as ‘ubercookies’ [6]).

We probably need to have a W3C policy that specifications cannot proceed beyond a certain stage without the working group undertaking a privacy review (similar to the IETF’s requirement of a security considerations section in any draft).

## 6 Looking at Do Not Track

The W3C has a member submission on the subject of “do not track” [1]. This technology is interesting: it has a clear emotive appeal connecting it with “do not call” [7]. However, there are some obvious differences: if someone violates “do not call”, both the definition of violation and awareness of violation are obvious (a telephone call is made); if some web service ‘tracks’ me, there may be disagreement over what constitutes tracking, and I may well be unaware of it.

In addition, it is clearly only a technique for consenting web services. It is akin to hanging a “privacy please” door hanger on an unlocked door – most will respect it, but the persistent will simply walk in.

There is, therefore, an urgent need to document what, fairly exactly, it means. What stops working? If nothing stops working, from the user’s point of view, there is a risk that it will be turned on all the time. Can I login? Buy something? What constitutes ‘track’? If someone buys something, I can obviously record the purchase, and pretty clearly the affect on my inventory. Am I allowed to record statistical data (e.g. the type of goods bought at different times of day)? At what point does this ‘personally derived data’ turn into ‘tracking’?

There is a minor point to be made about the header: if we imagine that a privacy conversation, mediated through HTTP headers, between users and servers, will be useful, we might prefer to use a more general header name (e.g. “privacy” rather than DNT) and more mnemonic values (e.g. “privacy: do-not-track” rather than “DNT: 1”). However, these are just protocol strings, and we can always say “DNT is used as a general privacy header, and 1 means do-not-track” – it is just that we’d probably prefer not to end up doing this.

There is also the possibility of a response from the server. This also would need definition, and careful balance of incentives. What effect may it have at the user-agent? If it’s only ever used to criticize, for example (“you responded saying you were doing X, and I don’t think you were”) there is little reason to use it. Similarly, if it is normally invisible to the user, why would it be sent?

## 7 Looking at the Exclusion List

The member submission also has a proposal for an exclusion list. I have doubts about the efficacy of this, if it were widely deployed. Sites whose business model depends on their users seeing advertisements, for example, would probably object if it became commonly easy to view the site with the advertisements missing. Since the technique is, in a sense, ‘hostile’, they may feel no compunction in taking counter-measures; rapid cycling of their DNS registrations, for example. This technique looks likely to lead to an arms race, and in arms races, there are usually no winners.

## 8 Conclusions

There seems to be low-hanging fruit here – some fairly readily available options:

- Define do-not-track, and what it means in request and response;
- Define privacy terms and policy fragments;
- Own the privacy implications of implementing W3C specifications ‘naïvely’.

## 9 References

- [1] “Web Tracking Protection”, Microsoft W3C Member Submission, <http://www.w3.org/Submission/2011/SUBM-web-tracking-protection-20110224/>
- [2] “Do Not Track: A Universal Third-Party Web Tracking Opt Out”, J. Mayer et al., <http://tools.ietf.org/html/draft-mayer-do-not-track-00>
- [3] “Privacy Considerations for Internet Protocols”, B. Aboba et al., <http://tools.ietf.org/html/draft-morris-privacy-considerations-03>
- [4] “Cyberspace security – Identity management—Baseline identity management terms and definitions”, ITU-T X.1252
- [5] “Terminology for Talking about Privacy by Data Minimization: Anonymity, Unlinkability, Undetectability, Unobservability, Pseudonymity, and Identity Management”, M. Hansen, Ed., <http://tools.ietf.org/html/draft-hansen-privacy-terminology-02>
- [6] “Cookies, Supercookies and Ubercookies: Stealing the Identity of Web Visitors”, A. Narayanan, <http://33bits.org/2010/02/18/cookies-supercookies-and-ubercookies-stealing-the-identity-of-web-visitors/>
- [7] “FTC Consumer Alert – The National Do Not Call Registry”, <http://www.ftc.gov/bcp/edu/pubs/consumer/alerts/alt107.shtm>
- [8] “Privacy Icons: Alpha Release”, Aza Raskin, <http://www.azarask.in/blog/post/privacy-icons/>
- [9] “Preventing attacks on a user's history through CSS :visited selectors”, D. Baron, <http://dbaron.org/mozilla/visited-privacy>

## **Privacy vs. Personalization Paradox in Online Advertising**

W3C Workshop on Web Tracking and User Privacy

Paul Trevithick, Azigo

ptrevithick@gmail.com

The conventional wisdom seems to be that the online advertising industry is caught in an insoluble privacy vs. personalization paradox. On the one hand, behavioral tracking using third party cookies violates commonly accepted notions of privacy such as notice, consent and transparency, yet on the other, if tracking cookies were to be widely blocked through technical means or their use made illegal, ad revenues would drop dramatically and the free Internet as we know it would be threatened. We who have been working in the identity space, however, do not believe that privacy and personalization are dichotomous rivals that need be traded one for the other. We believe that an alternative path is possible.

As has been widely reported, the FTC is exerting pressure with its Do-Not-Track proposals and the Obama administration has recently called for legislation in this area. Senator John Kerry, with support from Senator John McCain, is proposing a privacy bill of rights that would severely limit the use of tracking cookies. The pressure on the industry to respond has become acute but the response seems to be one that serves neither privacy nor personalization fully – that is, the proposed solutions seem intent on providing consumers with the ability to opt-out of behavioral tracking using rudimentary cookie blocking tools while hoping that too few consumers will actually use these tools in practice to affect the status quo: a mostly free Internet supported by advertising powered by revenue-enhancing yet invasive behavioral tracking of individuals by hundreds of third parties.

But, there is another path forward. The paradox can be resolved so that people can enjoy and benefit from the personalization of their web experience without sacrificing privacy, by enhancing their browsers with “self-tracking” capabilities under their control. By providing user-controlled, open standards-based “personal data services” through add-ons to existing browsers, the users themselves can curate their own rich behaviorally-derived profiles, hold them securely, and make profiles or relevant parts thereof selectively available to trusted websites for content and ad personalization. Where individual users do not wish to manage their own profiles directly, they can easily delegate management to trusted third parties without requiring non-consensual third party data aggregation, databases, or persistent cookies.

Our firm, Azigo, is developing a “data wallet” browser extension and personal cloud data service that implements this approach. It provides a privacy-aware foundation that also enables robust personalization.



March 25, 2011

Lorrie Faith Cranor  
Carnegie Mellon University  
Thomas Roessler  
World Wide Web Consortium

**Re: Proposal for Browser Based Do-Not-Track Functionality**

Dear Dr. Cranor and Mr. Roessler:

Thank you for the opportunity to submit this proposal regarding browser based Do-Not-Track. As the W3C continues to explore how best to build standards in this important area, BlueKai encourages the W3C to focus on increasing transparency and recognizing the economic tradeoffs that are a critical component to any discussion of Do-Not-Track. Unfortunately, in all too many discussions around the topic of online tracking, the economic tradeoff of free content for tracking is rarely discussed. The goal for this proposal is simply to establish a mechanism whereby individual website publishers will have the ability to highlight that economic tradeoff to consumers in a fully transparent way.

**Background on BlueKai and Transparency**

BlueKai ([www.bluekai.com](http://www.bluekai.com)) recognizes the importance of transparency, and indeed is the first online data exchange designed with consumer transparency and control in mind. We promote polite marketing on the Internet. BlueKai provides innovative services that help websites efficiently gain access to the vital advertising revenue on which they rely to offer services and content to consumers at a low cost. BlueKai partners with websites to collect “preference data,” which is anonymous information about visitors’ behaviors and interests. BlueKai also offers tools designed to enable consumers to see and change the preference data stored for a specific computer, or to opt out of the system entirely. For those computers where consumers have not opted out, BlueKai acts as an intermediary between the websites and advertisers (and ad networks) by providing pricing, standardization, and quality control for disclosing this anonymous preference data for advertising purposes. In this role, BlueKai contracts with websites to help them decide which advertising partners may target data gathered from websites and consumers, and also works with ad networks to help them locate trusted and high-quality data sources. BlueKai also controls the scope of information collected so that sites can protect themselves from unwanted data capture. Additionally, BlueKai sets terms and conditions and limits on how that data may be used. These features all underscore BlueKai’s commitment to transparency and privacy by design.

BlueKai is one of those businesses that, while not well known to consumers, plays a key role in maximizing consumer preference data for the delivery of relevant advertisements on publisher sites. BlueKai proudly offers an effective preference manager tool for consumers—the BlueKai Registry (<http://tags.bluekai.com/registry>).



The BlueKai Registry gives consumers complete access to the anonymous preferences associated with their computers that we store. Consumers can change the interest categories of preferences we record or remove preferences that do not interest them, thus declining to have their preferences used for advertising purposes. This Registry provides consumers with one place to impact many creators of data and many users of data as BlueKai works with several key data creators. We only share preference data pertaining to shopping or reading interests and do not collect data that consumers may consider to be sensitive, such as data on health, political interests, or adult behavior.

To help others provide the transparency that BlueKai has incorporated into our offerings, BlueKai launched a free white-label version of our BlueKai Registry in June 2010. This out-of-the-box tool for publishers and marketers gives consumers the same transparency and access to profiles through registries on their own websites at no charge to the companies.

### **Economic Tradeoffs and Do-Not-Track**

Advertising has always had two directions: interruptive or relevant. TV advertising is interruptive (it forces you to watch the ad and not the content through a commercial break) while search advertising is relevant (you are not forced to view the ad and the ad is so useful it is considered content). Data targeting is the fundamental technique by which online advertising becomes relevant and therefore, a publisher can show fewer ads to achieve the same revenue. Without data targeting, publishers can either force users to pay, or force them to see the ad before the content. (Or both). Polls of users such as that done by MarketingSherpa have show that overwhelmingly users (even the ones that don't like ads) prefer to get free content sponsored by targeting OVER having to pay for the content.

Therefore, we strongly encourage the W3C to ensure that any DNT functionality provides the marketplace with the opportunity to recognize the full economic tradeoff that consumers are making when it comes to online tracking.

### **The BlueKai Proposal for Do-Not-Track**

With that in mind, BlueKai is building a tool that will enable publishers to: a) recognize the Do-Not-Track header message, b) incorporate such header information into that publisher's content delivery process and c) provide consumers with the ability to disable or override Do-Not-Track in order to gain access to particular content.

### **Enable website publishers to recognize the Do-Not-Track header message**

The tool will enable website publishers to easily recognize a Do-Not-Track header message so that the website is in position to take action in response to the DNT header message.





**Enable website publisher to incorporate the Do-Not-Track header message into their larger content management process**

Once the Do-Not-Track header message is recognized, the tool will help the website integrate the DNT header into its content management process. Thus, the tool enables website publishers to establish content tiers such as free and premium in much the same way as some websites offer certain content to subscribers and non-subscribers. The Do-Not-Track header message would be available to publishers to be utilized as an input in its content management processes.

**Enable website publishers to provide consumers with the ability to disable or override Do-Not-Track in order to gain access to particular content.**

The tool will enable Individual websites to ask consumers to provide their consent to domain-based exemptions from the DNT header that include all third parties operating on a particular domain. In other words, the tool allows a consumer visiting a particular domain to enable ALL tracking for that domain.

For example, a consumer visits XYZ.com with the DNT header turned 'on'. XYZ.com asks the consumer to turn off DNT for that domain (e.g., in order to access premium content). Key to this approach is that the exemption must apply to all third parties (networks / platforms / exchanges) operating on that particular domain.

**Summary**

BlueKai believes that any discussion of Do-Not-Track should reflect both transparency and informed choice whereby consumers are informed of the full economic consequences of their privacy choices. And the BlueKai tool will establish a mechanism whereby individual website publishers will have the ability to highlight the economic tradeoff around Do-Not-Track to consumers in a fully transparent way.

Thanks again for the opportunity to present our proposal. We look forward to discussing this with you in the upcoming months.

Sincerely,

Omar Tawakol  
Chief Executive Officer  
BlueKai



## **CASRO AND ESOMAR POSITION PAPER**

*submitted March 24, 2011 for the*

### **W3C ONLINE TRACKING AND USER PRIVACY WORKSHOP**

#### **Introduction**

The Council of American Survey Research Organizations, Inc. ("CASRO") and ESOMAR are pleased to submit this position paper and declare our interest in attending the W3C Workshop on Web Tracking and User Privacy.

As the foremost research trade association in the United States, CASRO has long championed the public's right to privacy. CASRO is a not-for-profit association representing nearly three 350 research companies engaged in opinion, social, and marketing research regarding a wide variety of public and private issues.

ESOMAR is the essential organization for encouraging, advancing and elevating market research worldwide. With more than 4,800 members from over 120 countries on both the provider and client side, as well as in public bodies and academic institutions, ESOMAR's aim is to promote the value of market and opinion research in illuminating real issues and bringing about effective decision-making. Together with other industry associations, ESOMAR is representing the sector to the European Commission and Council of Europe, and is working closely with CASRO which represents the sector to the FTC, taking into account the need for a harmonized global perspective relating to online regulation.

Both CASRO and ESOMAR actively advocate responsible and ethical conduct through self-regulation. The [CASRO Code of Standards and Ethics for Survey Research](#) and the [ICC/ESOMAR International Code on Market and Social Research](#) set forth principles that guide our professional activities, such as requiring researchers to respect and protect the privacy of individuals who participate, whether passively or actively, in social, opinion and marketing research. Core to such self-regulatory Codes is that personal data collected for research purposes must not be used for other purposes and consent must be obtained if further processing is intended at a later date.

#### **Our position**

We support do not track proposals to the extent that they allow consumers to opt-out of online behavioral advertising (OBA). The scope of do-not-track should be limited to this activity and tracking where there is criminal or malicious intent. Website analytics and tracking activities conducted by research organizations for legitimate research purposes should be excluded.

Accordingly, we support the HTTP header approach in browser-based do-not-track tools, which would signal to advertising networks that users do not want their online activities tracked across multiple websites for advertising and marketing purposes. We hope that advertisers and marketers honor these requests.

We also support browser do-not-track tools that allow users to better manage their cookies, both browser cookies and local shared objects. Reputable advertising networks offer opt-out cookies to help recognize users who choose not to receive behaviorally-targeted ads. When consumers use their browser privacy settings to remove all cookies, though, their desired opt-out cookies could also be deleted. Add-ons are available for browsers to permit users to persist opt-out cookies, which we support.

We are concerned, though, that another browser-based do-not-track approach, that of using filter lists to block content and tracking scripts, could extend beyond OBA to include legitimate website analytics and research activities. Filter lists that include research trackers could have unintended and undesirable consequences for online panel research firms that obtain explicit consent from individuals to monitor their online behaviors.

Worryingly for consumers, we think that filter lists also have the potential to cause them harm. Conceivably, consumers could download filter lists from a website that they believe is reputable and trustworthy, but this turns out not to be the case. It is possible that consumers could unwittingly download what they believe is a legitimate filter list from a trusted website, but which is in actual fact spyware from a spoofed website. Cybercriminals and identity thieves now have a new means by which to dupe and exploit consumers.

In another scenario, filter list users could visit a website where they currently receive valued content for free and be informed that in order to continue receiving free services, they must download the website operator's filter list. Conceivably, the site's filter list could be set to:

- i) allow third-party scripts from an advertising network with which the site does business; and
- ii) block third-party scripts from the advertising network's competitors.

Tracking lists could thus be used in unintended and unscrupulous ways that have little to do with protecting consumers' privacy. Indeed, harm could be caused to consumers and to the marketplace.

In addition, we anticipate that the filter list approach could result in an explosive growth in the number of domains involved in online tracking activities, as unscrupulous tracking companies seek to avoid detection by registering multiple domains. Filter lists could give rise to a "whack-a-mole" situation that could prove to be unwieldy. The laudable objective of giving consumers an effective means to filter and control OBA tracking would be severely tested.

Accordingly, we believe that the most effective way to protect consumers from harm and ensure that their do-not-track requests are respected is for the U.S. and other national governments to pass laws that would require companies to honor consumers' wishes not to have their online activities tracked by third parties for OBA purposes. We favor a header approach in which the defined scope is third-party tracking for OBA purposes and we support browser extensions that persist opt-out cookies.

If filter lists are used, though, we believe that regulatory oversight is necessary to define the scope of third-party tracking activities that should be covered in filter lists and which activities should be excluded.

We wish to note that the activities of reputable market, social and opinion researchers are different from marketing, selling and advertising activities. It is important to summarize the distinction.

### **Market, opinion and social research is distinct from marketing, selling and advertising**

Opinion, social, and marketing research is distinct and separate from marketing, sales, and advertising activities and should not be subject to regulations aimed at those activities. While research is used by marketers to test their product or messages, it is not a promotional communication.

Market research, which includes social and opinion research, is the systematic gathering and interpretation of information about individuals or organizations using the statistical and analytical methods and techniques of the applied social sciences to gain insight or support decision making. Research elicits opinions and gathers information on behaviors, attitudes, characteristics, and possessions; it does not solicit money or invite purchases.

Research serves a critically-important function throughout our society to support decision making and to achieve that function, it must, and does, hold to the highest ethical standards of social science inquiry. It is utilized by universities, corporations, research institutes, litigants, politicians, and government agencies to develop behavioral and attitudinal data in support of technical, scientific, economic, health care, pharmaceutical, commercial, social and public policy issues. No other tool permits these constituencies to obtain comparable data or insights capable of serving as a barometer of public sentiment, behaviors, needs and aspirations. Without research, many issues affecting both public and private interests could not be addressed as intelligently or resolved as effectively.

It is important to note that the point of research is not to collect identifiable information for direct action, but rather to measure the behavior of small samples of a defined population in order to ascertain the views or behaviors of the whole population from which the sample was drawn. The risk of harm or adverse consequences for respondents where research is conducted in accordance with professional practices and under the oversight and enforcement of industry codes is infinitesimal.

U.S. federal law has supported the distinction between opinion, social and marketing research and marketing, sales and advertising activities. The Federal Trade Commission acknowledged the importance of research throughout its recent report, *A Preliminary FTC Staff Report on Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers*. In addition, the FTC has previously written that research is “informational,” has “social utility,” and is “not commercial speech.” It has recognized that distinction by excluding research from regulations that are intended to cover sales, marketing, and advertising activities, such as the Telephone Consumer Protection Act, the Telemarketing Sales Rule, the National Do-Not-Call Registry, and the CAN-SPAM Act.

Accordingly, we believe that browser-based do-not-track tools and any online do-not-track regulations should respect the trust and goodwill that researchers have earned with the public.

The research industry, including opinion, social, and marketing research, must have the ability to access respondents in order to collect and analyze their opinions and behaviors. Research depends on

statistical techniques to improve the quality of the sample and representativeness is a key characteristic for research to be robust for evidence-based policy making. Do-not-track tools that block researchers' ability to access Internet users and measure their online behaviors could degrade the quality of the statistical information and insights that we provide and on which decision-makers in the private, public and not-for-profit sectors depend to better understand consumers, customers and citizens for economic efficiency, innovation and progress.

### **Online tracking for research purposes**

We wish to provide an example of how reputable online panel research firms implement behavioral tracking research among their panel members. From this example, it will be apparent that our industry stands to be adversely affected by do-not-track lists that include research domains.

Online research panels comprise individuals who have agreed to participate in online survey research. Prospective members join a panel by filling in an online form on the panel research firm's website. The form requests basic profile and demographic information and may include other data, such as hobbies and interests. The profile information is used by the panel research firm to select individuals who meet the eligibility criteria for a particular study, e.g. a government agency wishes to test an anti-smoking advertising campaign among young people, aged 18 to 25.

When prospective panel members submit their profile data online, they typically also agree to the terms of participation and the site's privacy policy. In exchange for receiving periodic survey opportunities by email, text message or via a mobile application, panel members receive points that can be redeemed for cash or merchandise when they complete online surveys.

Research firms use first-party cookies for site administration and survey quality control purposes and this is explained in the firms' privacy notices. Some panel research firms offer panel members *optional* cookies for tracking research purposes. If panel members elect to receive the optional cookies, it is with their full knowledge and consent, and they can opt-out at any time by logging into the panel website and indicating their preferences or by contacting the panel manager. Advertising or website content that a sponsoring company would like to measure contains a script that is designed to read the optional cookie that the panel member has explicitly agreed to receive.

Thus, if a research firm's domain is captured in tracking filter lists that are downloaded by panel members, the result is the firm's scripts embedded in ads or website content displayed on third party sites will be blocked. In many cases, the blocking will occur against the wishes of individuals who have agreed to participate in the research firm's optional behavioral tracking research program in exchange for additional survey opportunities (e.g. advertising recall research) that earn them rewards. Most consumers that download tracking filter lists are likely not going to take the time to read upwards of 4,000 or more domains to see if the research firm's domain is included.

From the research firm's perspective, fraud is another possibility. Some panel members could explicitly agree to participate in a research firm's behavioral tracking research program. They could then download a filter list that, with their full knowledge, blocks the research firm's domain.

Panel members, whether unintentionally or deliberately, could thus block research firms from reading the optional cookies that they agreed to receive. For their part, research firms would not know that

their scripts were being blocked. Research firms could thus pay rewards to consumers who are not holding up their end of the bargain.

The integrity of behavioral tracking research conducted by panel research firms, which rely on explicit consent, is thus undermined by filter lists that capture research domains.

## Conclusion

In our position paper, we have identified critical issues that need to be addressed and resolved in a timely manner, including:

- The scope of what should be covered in do-not-track browser-based requests;
- The need for do-not-track laws to define the scope;
- The need for careful consideration and regulatory oversight regarding the use of tracking filter lists so that they do not create unintended consequences and cause harm to consumers, fair competition, and the integrity of research;
- The need to recognize the value of research and the responsible conduct of market, opinion and social research organizations that follow established and recognized codes of conduct.

We look forward to the prospect of discussing these issues at the upcoming W3C Online Tracking and User Privacy Workshop.

Respectfully submitted on behalf of CASRO and ESOMAR,



Diane K. Bowers  
President  
CASRO

170 North Country Road, Suite 4  
Port Jefferson, New York  
Tel: +1 631-928-6954  
Fax: +1 631-928-6041  
casro@casro.org  
www.casro.org

ESOMAR

Eurocenter 2, 11th floor, Barbara Strozziilaan 384  
1083 HN Amsterdam, The Netherlands  
Tel: +31 20 664 2141  
Fax: +31 20 664 2922  
public.affairs@esomar.org  
www.esomar.org

# **Empowering Users to Express a “Do Not Track” Rule: A Step Toward Conveying User Privacy Preferences**

**W3C Web Tracking and User Privacy Workshop  
April 28-29, 2011**

**John Morris and Alissa Cooper  
Center for Democracy & Technology**

## **1. Introduction**

This paper considers the privacy protection theory that underlies the “do not track” header and DOM property proposals, and addresses (and tries to rebut) a number of arguments that have been advanced against similar proposals considered elsewhere within the W3C.<sup>1</sup> Fundamentally, the idea of these do not track proposals is that a user-set privacy rule – an instruction not to track the user’s browsing – is included as a header in HTTP requests transmitted to web servers and/or as a property in the DOM. These approaches are instances of a broader approach to privacy protection – that of conveying user privacy preferences to entities that are in position to act on them

We believe that conveying user privacy preferences in general – and the do not track header and DOM property proposals in particular – have strong potential to be of value in the effort to protect privacy. There are other possible approaches to achieve a do not track regime, as summarized in a companion submitted paper entitled “Summary Comparison of Universal Opt-Out Mechanisms for Web Tracking” (and as detailed more fully in an Internet Draft recently submitted to the IETF<sup>2</sup>). The authors of both submitted papers are supportive of all of the approaches, and indeed we believe that a number of complimentary approaches could be implemented. We focus on the header and DOM property here for the purposes of drawing comparisons to previous efforts in the W3C, and urging further consideration of more broadly allowing users to set and convey their privacy rules.

The general approach of conveying user privacy preferences has been considered in at least three separate contexts within Internet standards bodies. First, starting in 2001, the Geopriv Working Group at the IETF implemented this approach by attaching privacy rules to location data.<sup>3</sup> Second, the Geolocation WG of the W3C considered a similar approach,

---

<sup>1</sup> Many of the points made in this paper were first articulated in a paper, “Binding Privacy Rules to Data: Empowering Users on the Web,” submitted by the authors (and Erica Newland) to the W3C Privacy Workshop held in July, 2010. That paper focused more broadly than just on “do not track” proposals, which are instances of the broader concept of allowing users to set and transmit rules to restrict third party use of their information and activities.

<sup>2</sup> See “Overview of Universal Opt-Out Mechanisms for Web Tracking,” available at <http://tools.ietf.org/html/draft-cooper-web-tracking-opt-outs-00>.

<sup>3</sup> See <http://datatracker.ietf.org/wg/geopriv/charter/>. One of the authors of this paper, Alissa Cooper, is a co-chair of the Geopriv WG, and the other, John Morris, is a co-author of a number of Geopriv RFCs.

but rejected it for many of the reasons discussed (and rebutted) in this paper.<sup>4</sup> Third, the Device API and Policy Working Group (DAP) of the W3C has explored the notion of passing user “privacy rulesets” to consumers of the DAP APIs,<sup>5</sup> but the proposal has met resistance from some WG participants for many of the reasons discussed in this paper.<sup>6</sup> This paper briefly addresses the criticisms raised by opponents to the approach of conveying user privacy preferences, and assesses those criticisms in the context of the do not track header and DOM property proposals.

## 2. Conveying User Privacy Preferences

The central feature of conveying user privacy preferences is that when a user communicates with another entity, applicable privacy rules are also conveyed to the entity to ensure that entities that receive information about the user are informed of how they may (or may not) use it. By creating a structure to convey the users' preferences along with their information or communications, the likelihood that those preferences will be honored necessarily increases. In particular, no entity can disavow knowledge of users' preferences for how their information may or may not be used. Conveying user privacy preferences allows users to express their desire for and expectations of privacy, which in turn helps to bolster social and legal systems' protection of those expectations.

Applying and affixing usage rules to information is a well-known way of protecting information, long before the World Wide Web (for example, by placing the © copyright symbol on documents). More recently, the Creative Commons<sup>7</sup> model is one prominent example, allowing an owner of a work to set four types of rules ("Attribution," "Noncommercial," "No Derivative Works" and "ShareAlike") governing the subsequent use of the work. After the author sets these rules, the rules are conveyed together with the work itself, so that every consumer of the work is aware of the copyright terms.

Another example where usage rules are bound to data is in security classification systems (such as marking documents with a “Secret” designation). As these examples reveal, these systems of rule enforcement are *not* self-executing. Unlike some technical strategies (such as encryption), these systems rely on external, *non-technical* mechanisms (such as laws, contracts, or company rules) to enforce the protection of the information. The do not track header and DOM property proposals follow this model – they propose the creation of a technical requirement to ensure that the applicable user preference is always conveyed to entities capable of tracking, and it leaves to regulatory, legal, and market forces the enforcement of the user’s directive.

---

<sup>4</sup> See <http://www.w3.org/2008/geolocation/>.

<sup>5</sup> See <http://dev.w3.org/2009/dap/privacy-rulesets/>.

<sup>6</sup> See <http://www.w3.org/2009/dap/>. The authors of this paper have been among the primary advocates for the approach of binding user-set rules to data within both the GeoLocation and DAP working groups.

<sup>7</sup> See <http://creativecommons.org/>.



### 3. Arguments Against Conveying User Privacy Preferences

In the Geolocation WG and, to a lesser extent, the DAP WG, a number of arguments have been raised against the idea of conveying user privacy preferences. This section briefly recaps some of the criticisms and responses to them, without intending to be an exhaustive discussion of either side of the arguments.

#### **a. Conveying user privacy preferences does not protect privacy through technical means (such as encryption).**

Conveying preferences does not, by itself, provide technical means through which it can be reasonably guaranteed that users' privacy rules will be honored by recipients of their data. Thus, the transmission of a do not track header, for example, does not in any technical way assure that the recipient web site receiving the header will not, in fact, track the user. Instead, the privacy protection is provided by virtue of the fact that data recipients are informed of the user's preference, and are expected to only use data in accordance with that preference.

By conveying the user's preference, the approach provides valuable information so that *non-technical* forces such as legal contracts, governmental consumer protection authorities, and marketplace feedback can better enforce those preferences. If a commercial recipient violates a user's clear privacy preference, for example, the recipient can, in a growing number of countries, be charged with violating consumer or data protection laws. In the absence of an expressed preference, consumer protection authorities are less able to protect consumers whose information has been abused.

#### **b. Implementing a preference interface in a user agent would be hard, and users might be confused.**

Without question user interfaces are hard. But given that the user agent serves as a crucial gateway between users and the web, providing centralized privacy preference interfaces in the user agent may in fact help to reduce the confusion caused by each individual web site or app giving users different controls and interfaces over essentially the same user data being communicated through the user agent. User agents already contain privacy preference interfaces, for example to control cookies. When cookies were first introduced on the web, browsers provided no way for users to control their use.<sup>8</sup> As concerns were raised about potentially privacy-invasive uses of cookies, browser vendors began to add cookie controls into their products, beginning with rudimentary tools and evolving over time to the more sophisticated controls in place today. Browser makers continue to explore simple ways to present privacy choices in the browser,<sup>9</sup> and interfaces to convey user preferences should be part of that exploration.

---

<sup>8</sup> See *Federal Trade Commission Staff Report. Public Workshop on Consumer Privacy on the Global Information Infrastructure, Part III: Enhancing Consumer Protection Online* (Dec. 1996), available at <http://www.ftc.gov/reports/privacy/Privacy4.shtm>.

<sup>9</sup> See, e.g., <http://people.mozilla.com/~faaborg/files/firefox4Mockups/prefPaneWebSites-i2.png>.

**c. Users would blame the browser when web sites violate the users' expressed preferences.**

If a browser provides a user interface allowing users to set a privacy preference (such as do not track) and those rules are later violated, there is a risk that the browser will be blamed. There are, however, affirmative steps that can be taken when designing the user interface to mitigate this possibility. A user interface can make clear that it is soliciting preferences to be conveyed to the recipient of the information, and that the recipient is responsible for honoring them. The user interface associated with the do not track header in Firefox 4 provides a good example: users can check a box to "Tell web sites I do not want to be tracked," as opposed to a box that says "Do not let web sites track me."<sup>10</sup> By being careful to convey the limits of the browsers' control over later tracking or other uses of the users' data and providing supplemental user education about the user agent privacy settings,<sup>11</sup> the user agent can reduce the risk that it would be blamed for a privacy violation by a receiving entity.

**d. Rather than providing incomplete privacy protection, it is better for users to think there is no privacy protection.**

In the security context, there may be real risk if users mistake weak protection for adequate protection – they may expose critical data (such as, say, bank account login information) and then suffer catastrophic harm. And there often is an available way to achieve real security, even if it means a delay or inconvenience in performing a transaction.

The privacy context is quite different. The harm is often more incremental, and users are better off if even a subset of recipients of their information or communications honor their privacy preferences. In the web tracking context, users may not fully understand the extent to which data about their web behavior is being collected and used, but they may still want to convey their preference not to be tracked, even if it is not universally honored from the outset. In the context of web applications that require users to affirmatively share data about themselves, users are often presented with a "Hobson's choice" with regards to their data: using a service requires implicit acceptance of all future data uses by the service provider, and the only other option is to not use the service at all. Unlike in the security context, users often have no alternative to this "take-it-or-leave-it" approach to privacy, and so users are forced to give up their privacy. Any enhanced privacy protections, even if incomplete, will offer users a substantive improvement over the status quo.

**f. We are not sure it will work.**

The approach of sending user privacy preferences is new to the applications layer, and there is certainly no guarantee that this framework will succeed. But, one thing is certain:

---

<sup>10</sup> See <http://support.mozilla.com/en-US/kb/how-do-i-stop-websites-tracking-me>.

<sup>11</sup> See, e.g., <http://ie.microsoft.com/testdrive/Browser/TrackingProtection/Default.html>, <http://support.mozilla.com/en-US/kb/how-do-i-stop-websites-tracking-me>.

the status quo has failed to provide meaningful privacy protection on the web. Privacy policies are not widely read or understood<sup>12</sup> while web tracking continues to become more sophisticated and pervasive<sup>13</sup> despite users' discomfort with it.<sup>14</sup> Doing nothing to change this situation in the face of the growth of web applications will only further jeopardize user privacy on the web.<sup>15</sup>

\* \* \* \*

By using a do not track header and/or DOM property, users can be given some element of control, as well as some legal claim, over whether their web browsing is tracked.<sup>16</sup> The same argument could be made for building mechanisms that would allow users to express their privacy preferences over geolocation and other types of sensitive personal information. The work of some browser vendors to implement a do not track header and/or DOM property is strong evidence that the objections to the approach of conveying user privacy preferences are in fact surmountable. As with do not track, placing users in the position of being able to set – and have expressed in a standardized way – their privacy preferences will greatly increase the chance that those preferences are honored.

---

<sup>12</sup> See, e.g., <http://portal.acm.org/citation.cfm?id=1614511>.

<sup>13</sup> See, e.g., <http://web.cs.wpi.edu/~cew/papers/soups07.pdf>.

<sup>14</sup> See, e.g., [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1478214](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214).

<sup>15</sup> Significantly, legislators and regulators in the United States are increasingly interested in the potential that do not track techniques offer. Although the U.S. Federal Trade Commission has authority to regulate unfair and deceptive trade practices – and could pursue violations of do not track instructions based on its current authority – some in Congress are considering proposals to make explicit the FTC's authority to enforce do not track rules set by users.

<sup>16</sup> As noted above, the authors do not believe that a do not track header and/or DOM property are the only approaches to web tracking opt outs that are worth considering, and other approaches may also be useful tools to provide to users.

# W3C Workshop on Web Tracking and User Privacy

## Position Paper

### Online Tracking, Targeting and Profiling: A Canadian Privacy Perspective

**Andrew S. Patrick**  
IT Research Analyst  
Office of the Privacy Commissioner of Canada

The paper is based on the *Draft Report on the 2010 Office of the Privacy Commissioner of Canada's Consultations on Online Tracking, Profiling and Targeting and Cloud Computing* (October 25, 2010). The full report can be found at [http://priv.gc.ca/resource/consultations/report\\_2010\\_e.cfm](http://priv.gc.ca/resource/consultations/report_2010_e.cfm) and the final report will be issued in May, 2011.

In the spring of 2010, the Office of the Privacy Commissioner of Canada (OPC) held consultations on online tracking, profiling and targeting. The OPC received 21 written submissions and held two public events in Toronto and Montreal attended by representatives of industry, as well as academics, advocates, and members of the public.

The written submissions focused primarily on behavioural advertising — what it is, what the benefits and risks are, and what self-regulatory measures are in place. Many respondents and participants raised various privacy issues including the blurring of the public/private divide and its effects on reputation was seen as a significant issue. Children's activities online and the need to incorporate privacy into digital citizenship programs were also concerns that were raised.

The OPC believes that traditional notions of public and private spaces are changing. Canadians, though, continue to consider privacy to be important but they also want to engage in the online world. The two are not mutually exclusive, but we think more needs to be done to protect privacy so that individuals can trust those offering her products, services and places to be social.

It is still early days in terms of research into people's perceptions of their audience and the possible disconnect between who they think their audience is and the reality. Complicating how people communicate and interact online, as researcher danah boyd notes, is that social networks, in particular, have certain properties that alter social dynamics: persistence, searchability, exact copyability, and invisible audiences. In terms of social networking activity, some early research suggests that individuals do make distinctions with respect to their intended audience and wish to exert some measure of control. The difficulty in exerting control lies in the architecture of a site. When privacy controls are difficult to find or understand on a web site, the ability of the individual

to exert any control drops. If the site is popular and the individual is keen to be part of the community, he or she may risk being more open in order to participate in the site.

The OPC questions the view that since people put information "out there" that it therefore is available for any kind of use. Some research is showing that people intentionally project specific personas online and post information that will support these personas, usually to gain some status. It is not clear that the intention is always to be public. For example, someone may want to cultivate a professional presence online, but they may also want a separate social space to engage with friends outside of the work context. Making and keeping these worlds separate is neither obvious nor easy.

Moreover, in Canada, although personal information may appear in the public domain that does not necessarily mean it can be used for any purpose. For example, PIPEDA (Canada's private sector privacy law) provides that some publicly available personal information can be collected, used and disclosed without an individual's consent; however, the purposes for which that information may be collected, used or disclosed, are nonetheless limited.

The OPC is of the view that the consequences of the apparent breakdown between public and private lives can be seen most clearly in terms of harm to real world reputations. Individuals — teachers, politicians, police officials — have lost jobs, been publicly embarrassed, or lost benefits because of what they have posted online. Online, data persists. Information that harms an individual's reputation may never really go away. Moreover, with the increasing popularity of location-based applications, one consequence of telling people where you are is that you also tell them where you are not, potentially leaving one's home at risk.

There are also implications with respect to the accuracy of the profiles data miners construct. Much has been made about the use of social network profiles in determining employability or in determining acceptance to post-secondary education facilities. However, tracking and profiling online browsing behaviour also has consequences and is of great concern given the near invisibility of the practice. If these practices only resulted in targeted marketing, the risks of inaccuracy might seem minimal (although it could be problematic if people do not receive benefits that others do). If profiles are used more broadly, perhaps for granting loans, assessing insurance risks or assessing national security risk, the unforeseen consequences can be potentially more serious. There are also other potentially serious public policy issues that do not touch on privacy, such as limitations on freedom of speech.

The concept of "harm" appears to be used by some to distinguish certain practices that should require consent and those that should not. It should be noted, however, that PIPEDA does not contain such a concept. Rather, it requires that purposes be "appropriate", identified to the individual and consent obtained (the type of consent may vary). Instances where consent is not required are limited.

The OPC has been following developments in the area of identity management as part of its strategic priorities. Identity management may be helpful in providing individuals with better means of controlling their personal information but it also has privacy implications in that, if not done well, it may make it easier for data to be linked to previously separate identities. We are interested in the ideas surrounding "digital identity" being proposed by Kim Cameron and others. Digital identities should be flexible so that they sometimes correspond with natural, flesh-and-blood identities, and sometimes they are completely separate. Identities should allow someone to be public and private, according to the context. Identities should also allow the verification of a claim (e.g., old enough to drink) while adhering to a principle of minimal disclosure (e.g., not revealing the actual date of birth). We are tracking efforts to develop identity metasystems that allow for the effective creation and management of different identities.

The OPC supports the view that privacy considerations should be a critical component of the design stage of any technology or use of technology. In our recent submission to the Government of Canada on the Digital Economy Strategy, we noted that more could be done to prevent privacy problems or mitigate the effects on privacy protection posed by new technology by making privacy an integral part of the development of the digital economy. Other data protection authorities in other parts of Canada and the world are calling for "privacy by design" to be required in data protection legislation. The Information and Privacy Commissioner of Ontario, Ann Cavoukian, has been a long-time proponent of the concept of privacy by design.

The OPC is also of the view that privacy needs to also become an integral part of the business processes and models that rely on technology through a careful analysis of companies' activities. Privacy impact assessments (PIAs) are a useful tool that the private sector should be encouraged to use, since greater emphasis on such analysis may prevent problems from arising.

Expecting users of the web to navigate the privacy implications of the many services and business practices online, understand these implications, and consent to the practices may be unreasonable without a strong baseline of privacy protection. Knowledge and consent are key in PIPEDA but there are other principles that organizations need to consider more carefully and build into technology and business models.

## Comcast Position Paper for Submission to the W3C Workshop on Web Tracking and User Privacy<sup>1</sup>

Protecting the privacy of consumer information transported over the Internet deserves the high-priority attention of all stakeholders in the emerging marketplace of online communications and commerce. Consumers are rightly concerned that the personal information they provide over the Internet may be collated, gathered, tracked and distributed in myriad ways so that far too many persons and entities will know far too much about them. Cable operators and programmers not only understand these concerns but are committed to protecting the privacy of their customers. Comcast, as a member of the cable industry, has actively participated in the privacy policy discussion with federal regulatory agencies, legislators, industry groups, and public interest groups. For Comcast, the upcoming W3C Workshop provides a critical opportunity to work with stakeholders and continue to advance the privacy policy conversation.

For cable operators like Comcast, the privacy of their customers is not a new concern. Since long before they began offering broadband service, cable operators have been taking steps to protect customers of their cable television service against any undesired disclosure of their personally identifiable information (“PII”) and their purchasing and viewing decisions. Since 1984, such measures have been required by federal law. But they’re also a business imperative – especially in today’s competitive broadband marketplace. For all the services that cable operators now offer – video, broadband and telephone – consumers have choices. Moreover, more and more consumers are now purchasing all these services from a single provider, so that the costs of losing a customer to a competitive provider are compounded. In other words, cable operators have singularly strong incentives to meet the privacy concerns and demands of their customers.

But *how* to meet the privacy concerns and demands of consumers when they use the Internet is a much more complex task, and it involves a much larger ecosystem of entities, many of which may not have the same ongoing relationship with – and incentive to protect – consumers’ privacy. Moreover, balancing those privacy needs against the uses of consumer information to support legitimate and beneficial Internet services and applications presents new and challenging issues for service providers and policymakers alike.

The Federal Trade Commission’s (“Commission”) Staff Report, *Protecting Consumer Privacy in an Era of Rapid Change*<sup>2</sup> (“Staff Report”) is a commendable effort to address those issues and help meet those challenges. The Staff Report provides a comprehensive analysis of the current state of privacy protection that identifies what appears to be working and what appears not to be working in ensuring that consumers’ interests are protected. It proposes a new “framework” for addressing Internet privacy concerns, setting forth its proposed framework as a

---

<sup>1</sup> This document is substantially similar to the “Introduction and Summary” of comments submitted by the National Cable and Telecommunications Association (“NCTA”), in response to the Federal Trade Commission’s Staff Report entitled *Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers* in February 2011. Though NCTA submitted this paper in its own name, Comcast is an active NCTA member and was a major contributor to this document. The paper accurately reflects Comcast’s position, which is echoed by our industry counterparts.

<sup>2</sup> Federal Trade Commission, *Protecting Consumer Privacy in an Era of Rapid Change*, Staff Report (Dec. 2010), available at <http://www.ftc.gov/os/2010/12/101201privacyreport.pdf> (“Staff Report”).

“policy vehicle,” leaving open the question whether the framework might effectively be adopted, in whole or in part, by the affected entities themselves, voluntarily or through self-regulatory mechanisms, or whether it must be mandated by the government.

Regulation in this area must be carefully developed so that it does not constrain the flexibility of Internet entities to tailor their privacy protections to changing technologies, new services, and the evolving economics of the Internet. Regulation – even self-regulation – virtually always produces unintended consequences. And the cost of unintended consequences is uniquely high when they could affect the enormously successful and beneficial Internet ecosystem. Self-regulatory mechanisms are worth exploring and developing in any event, because self-regulation can be quickly modified and adapted to remedy such consequences.

The Staff Report recognizes one of the ways that unduly restrictive or overbroad privacy requirements can have adverse consequences. Specifically, the Commission recognizes the importance of online advertising revenues to the economic underpinnings of Internet content and services. Such revenues supplement, and in many cases substitute for, fees that would otherwise have to be charged to consumers to support such content and services. Without them, the innovation, competition and constant expansion of available content and services that have been the hallmark of the Internet would be impaired. Moreover, forcing more of the Internet’s costs to be borne by consumers would undermine the public policy goal of encouraging greater availability and adoption of broadband services. One method of efficiently maximizing the availability of advertising revenues in such a highly competitive marketplace is so-called “targeted advertising” – advertising that is sent specifically to consumers who are most likely to be interested in particular products or services. Targeted advertising may implicate privacy concerns: How do advertisers identify the consumers who are most likely to be interested in their products? The benefits of such advertising must be balanced against such concerns in determining whether and to what extent it should be restricted.

The Staff Report includes many useful ideas and recommendations for balancing the interests at stake in developing a privacy policy framework. A pro-active policy of “privacy by design,” for example, minimizes the risk of privacy breaches and concerns from the outset and should be a fundamental component of the development of new Internet products and services by all responsible Internet companies. The cable industry, as noted above, is committed to protecting the privacy of its customers and, to this end, our companies are continually engaged in efforts to develop best practices and promote consumer privacy at every stage of the development of products and services (including the development of targeted advertising policies and procedures).

The concept of “notice and choice” should also play a role in any sound privacy policy insofar as it enables individual consumers to decide, in certain cases, whether the benefits of disclosure of certain consumer information in certain circumstances override any privacy concerns. But the effectiveness of notice and choice can be undermined if it is implemented in a way that is confusing to – or ignored by – consumers. The Commission’s proposal to simplify consumer privacy notices by removing from “notice and choice” those transfers of consumer information that are “commonly accepted practices” – or perhaps more appropriately, those for which there is no expectation of privacy – is a step in the right direction.



So, too, is the Commission's recognition that for those practices that remain subject to notice and choice, there may be no single best way to offer such notice and choice in all circumstances. Where disclosure of consumer information can provide benefits to consumers (such as in the case of targeted advertising), notice and choice should be designed to ensure that consumers understand both those benefits *and* the privacy implications. Reflexive opting *out* where a consumer does not fully understand and take into account the *benefits* of disclosure of information is as undesirable as reflexive opting *in* where the consumer does not understand or cannot be expected to take the time to read the details of how and when such information will be disclosed. In particular, a uniform "Do Not Track" button, while providing an easy way to opt out of a privacy-related practice, could lead to just this sort of reflexive and uninformed choice, with unintended and unwanted consequences for consumers. Figuring out how to adapt notice and choice to the vast array of different circumstances in which consumer information may be used and disclosed by Internet content, application, and service providers is precisely the sort of task best implemented through vigilant and ongoing self-regulation.

Caution is warranted before the Commission accepts the suggestion in the Staff Report that the distinction between PII and information that is not personally identifiable has been blurred to the extent that it should no longer be relevant for privacy purposes. Privacy policy (as embodied, for example, in the privacy legislation applicable to cable television operators) has until now generally recognized that the collection and disclosure of aggregate or anonymous data – which can serve wholly legitimate, beneficial, and pro-consumer purposes – does not raise the same concerns or require the same protections as the collection and disclosure of PII. There are also ongoing changes in privacy-enhancing "anonymization" technologies that are designed to *prevent* "re-identification."

Finally, there is a bedrock principle that appears to be missing from the Commission's otherwise comprehensive and commendable Staff Report – the principle of competitive neutrality. In the evolving Internet marketplace, competition extends across the multiplicity of categories of service providers. Cable operators compete, of course, with other broadband Internet Service Providers ("ISPs"), including telephone companies and, increasingly, wireless service providers. But ISPs also compete with other Internet entities – including entities with access to consumer information – in the highly competitive Internet advertising marketplace.

It is crucially important to a fair, efficient and well-functioning marketplace, as well as to the protection of consumers' privacy interests, that any privacy policies apply uniformly to particular *conduct* or types of data collection that affects the privacy interests of consumers and do not single out particular categories of service *providers* for special treatment. In particular, imposing unique or "heightened" restrictions on conduct simply because it is engaged in by broadband ISPs would be especially perverse. As discussed above, ISPs have unique incentives, because of their ongoing relationship with consumers and because of the high cost of losing a broadband customer to a competitor, to be *especially* vigilant in protecting their privacy.



## **Council of Europe Contribution to the W3C Workshop on Web Tracking and User Privacy**

28-29 April 2011, Princeton University

---

### **Convention 108: protection of individuals with regard to automatic processing of personal data**

The Council of Europe celebrates this year the 30<sup>th</sup> Anniversary of its Data Protection Convention (usually referred to as Convention 108) which has served as the backbone of international law in over 40 European countries and has influenced policy and legislation far beyond Europe's shores.

With new data protection challenges arising everyday, the Convention is being overhauled to meet new realities and the Council of Europe is currently working on its modernisation. If the principles of the Convention can be considered as time-proof, the latest technological developments of the information and communication society and the globalisation of exchanges nevertheless lead to potential new risks for the protection of human rights and fundamental freedoms, which may require specific attention.

In its modernisation work of Convention 108, the Council of Europe launched on Data Protection Day (28 January) a public consultation. A consultation paper identified several issues to discuss in the context of the modernisation and a number of interrogations and proposals were shared, one of them being specifically related to tracking as addressees were asked if a right 'not to be tracked' (RFID tags) should be introduced in the Convention. Responses to this consultation are now being considered and will be examined by the Consultative Committee of Convention 108 in the coming months.

Link to Convention 108:

<http://conventions.coe.int/Treaty/Commun/QueVoulezVous.asp?NT=108&CM=1&DF=&CL=ENG>

Link to the modernisation page:

[http://www.coe.int/t/dghl/standardsetting/dataprotection/Modernisation\\_en.asp](http://www.coe.int/t/dghl/standardsetting/dataprotection/Modernisation_en.asp)

## **Recommendation (2010)13 of the Council of Europe on the protection of individuals with regard to automatic processing of personal data in the context of profiling**

This Recommendation is the first text to lay down internationally-agreed minimum privacy standards to be implemented both by the public and private sector, through national legislation and self-regulation. It has been adopted by the Council of Europe, as a sectorial complement to Convention 108 on data protection.

Profiling, the technique of observing, collecting and matching people's personal data online, can benefit both individuals, the economy and society by, for instance, leading to better market segmentation or permitting an analysis of risks and fraud. However the use of profiling techniques without precautions and specific safeguards could severely damage human dignity by notably unjustifiably depriving individuals from accessing certain goods or services.

The Recommendation aims at:

- providing a coherent regulatory framework, which strikes a fair balance between the interests at stake;
- ensuring effective protection of the rights of data subjects and fair procedures in situations where mass quantities of data are processed;
- avoiding decisions, discrimination or stigmatisation made automatically on the basis of profiles.

Link to the Recommendation:

<https://wcd.coe.int/wcd/ViewDoc.jsp?id=1710949&Site=CM&BackColorInternet>

And to its explanatory memorandum:

[https://wcd.coe.int/wcd/ViewDoc.jsp?Ref=CM\(2010\)147&Language=lanEnglish&](https://wcd.coe.int/wcd/ViewDoc.jsp?Ref=CM(2010)147&Language=lanEnglish&)

### **Other relevant texts**

The Committee of Ministers of the Council of Europe adopted a Declaration on freedom of communication on the internet:

[http://www.coe.int/t/dghl/standardsetting/media/Doc/H-Inf\(2003\)007\\_en.pdf](http://www.coe.int/t/dghl/standardsetting/media/Doc/H-Inf(2003)007_en.pdf)

The Council of Europe also produced Human Rights Guidelines for Internet Service Providers:

[http://www.coe.int/t/information/society/documents/HRguidelines\\_ISP\\_en.pdf](http://www.coe.int/t/information/society/documents/HRguidelines_ISP_en.pdf)

### **What's next?**

The Council of Europe is currently preparing two draft Recommendations fully relevant to the workshop's topic: one on the protection of human rights with regard to search engines and another one on measures to protect and promote respect for human rights with regard to social networking services, which will both address the need for transparency concerning the use of personal data.

Both Recommendations will be addressed to member states/governments and complemented by guidelines aimed at providing inspiration/guidance to Industry.

# Location privacy in web-based LBS

## Position paper

Maria Luisa Damiani<sup>1</sup> and Pierluigi Perri<sup>2</sup>

<sup>1</sup> Dept. Informatics and Communication, University of Milan (I)

<sup>2</sup> School of Law, University of Milan(I)

## 1 Motivation and background

MODAP (Mobility, Data Mining and Privacy, 2009-2012) is a project funded by the European Commission to promote awareness of the privacy issues in mobility data collection and data mining (<http://www.modap.org>). The project consortium consists of 11 institutions from various European countries, for the major part universities. The University of Milan is one of the members of the project. Within our research, mainly focused on privacy-enhanced technologies [1, 3], we are experiencing interdisciplinary collaboration between jurists and researchers on the issue of location privacy in web-based location-based services (LBS)[2]. We call web-based LBSs those applications in which users can request a LBS through a geo-enabled browser compliant with the W3C geolocation API specification. Accordingly, the user visiting a geo-enabled website is prompted with the question on whether he/she gives consent to the disclosure of the location to the website owner. In our research, we are concerned with the analysis of the privacy risks emerging in this scenario and with the problem of how to enhance user's awareness for a more responsible user's participation. In this position paper we want to contribute to the discussion with some considerations.

## 2 Enhancing users' awareness

### 2.1 Who is tracking me?

Users are not fully aware of all parties, or Data Controllers in a privacy-oriented taxonomy, which track their position. For example the users accessing the Foursquare website through the Firefox browser deliberately decide to share their position with the members of the geo-social network and thus also with the website owner. It is very likely however that inexperienced users are not aware of the fact that, in doing so, they disclose their position to some third party (i.e., the location provider) other than the website owner. In essence, the location provider which computes the position on behalf of the geo-enabled browser is transparent to the user. Indeed the user has only evidence of the fact that the position is tracked by the website owner, without knowledge of how many other subjects may be included in this tracking. This follows from the compliance of browsers with the W3C Geolocation specification. For the sake of transparency the user should get the full information when the yes/no consensus is requested.

## 2.2 Freedom of choice

Location providers have the ability to track users with great precision across different (geo-enabled) websites. Moreover, it can be shown that personal and sensitive information can be easily extracted from the collected location data, e.g. the home address [2] or the hospital in which the user is undergoing a medical visit.

Now consider an user willing to share his/her position with a trusted website, say the website of the ecologist organization the user belongs to, without letting the location provider know that he/she is at home or hospitalized. Note that this scenario is specular to the one commonly adopted by the research community working on privacy in LBS [4] in which the LBS provider is untrusted while the location source is trusted. Of course one could say that disclosing the position to the location provider is the prize that the user must pay to access the LBS. Indeed, this is only in part true, because users are not allowed to choose the location provider, based on personal preferences. The only way for the user to interact with a different location provider is to use a different browser or operating system and that of course is not an usable solution.

## 2.3 No invasive privacy

Users, even those who are sensible to privacy, might desire not to be bothered by privacy when they are working or doing something else. On the other hand, specifying privacy settings by clicking on a set check-boxes can be extremely boring. Moreover if this operation is to be repeated for every geo-enabled application, it takes time and is costly. In this view it might be useful some form of automation working across multiple applications. In the simplest case, it could be a sort of "red button" that the user may activate to immediately, and easily, stop tracking. A more sophisticated solution would be trying to minimize the interaction with the location provider, to limit the dissemination of location information [2].

## 3 Conclusion

In summary, web-based LBSs offer extraordinary opportunities to location and LBS providers to collect huge amount of position data in a simple way. This also raises challenging opportunities of research on privacy enhanced technologies. Therefore it is important to bring this scenario to the attention of researchers working in the area, because the level of awareness seems still limited. In this perspective, experiments with the users of web-based LBS can be of vital importance to gain insights into user's expectation on privacy.

## References

1. M.L. Damiani, E. Bertino, and C. Silvestri. The PROBE Framework for the Personalized Cloaking of Private Locations. *Transactions on Data Privacy*, (3)2:123–148, 2010.

2. M.L. Damiani, P. Perri, and C. Yildizli. Third party positioning services: novel challenges in location privacy in LBS. Technical Report, March 2011. Universita degli Studi di Milano, TR38-11.
3. M.L. Damiani, C. Silvestri, and E. Bertino. Fine-grained cloaking of sensitive positions in location sharing applications. IEEE Pervasive Computing (accepted for publication, pre-print online).
4. M. Gruteser and D. Grunwald. Anonymous Usage of Location-Based Services Through Spatial and Temporal Cloaking. In *Proc. of the 1st International Conference on Mobile systems, Applications and Services*. ACM Press, 2003.



March 25, 2011

To the Program Committee:

Datran Media appreciates the opportunity to participate in W3C Workshop on Web Tracking and User Privacy. By way of background, Datran Media is a leading digital marketing technology company that helps advertisers and publishers discover and reach their ideal audiences. Datran Media provides digital advertising solutions, audience measurement and analytics, and marketing compliance solutions. Notably, PreferenceCentral – a service developed by Datran Media – is based on the goal of offering consumers meaningful choices in digital advertising. Specifically, PreferenceCentral enables companies to comply with the Industry’s Self Regulatory Principles while providing consumers the opportunity to make brand-level choices, instead of just ad network choices. As such, we offer this position paper for your consideration.

## **BACKGROUND.**

In May 2010, PreferenceCentral performed a survey of 1,050 Internet users. <sup>1/</sup> The survey asked consumers to state their preferences for tailored online advertising within a behavioral economic context of real world, value-for-value trade-off options. The survey revealed several interesting facts relevant here. First, Internet users are more likely to prefer targeted online ads when they are asked to make real-world, value-for-value tradeoffs, such as access to free content in exchange for targeted ads. Second, Internet users’ attitudes and preferences significantly shift when consumers are provided with education about behavioral targeting or when they are offered the ability to control targeted ad exposure. We believe that this survey demonstrates that consumers

---

<sup>1/</sup> Available at [www.preferencecentral.com/consumersurvey/](http://www.preferencecentral.com/consumersurvey/).



value targeted advertising, particularly when consumers receive valuable content in exchange. It also shows that this value is enhanced by offering consumers the ability to control their exposure to targeted ads.

Because of the importance of enabling these trade-offs for consumers, the goal of the Do-Not-Track mechanism should not be to block tracking, but rather to give the consumer meaningful choices about whether they want to be tracked and how the collected data can be used.

## **THE DO-NOT-TRACK HEADER**

Datran believes that a persistent browser-based mechanism such as the Do-Not-Track browser header is the right approach to implement Do-Not-Track. The benefit of this mechanism – rather than a tool that blocks tracking – is that consumers can communicate their preferences about not being tracked directly to website publishers and technologies. Any mechanism that implements Do-Not-Track should be required to include a simple notification and override (based on consumer opt-in) that can be used by web publishers and technologies. Essentially any Do-Not-Track mechanism should be a communication platform rather. A communication platform allows publishers and technologies to tailor an appropriate choice in response to consumers’ expressed preferences. For example, a website has the ability to offer free access to content if a consumer agrees to opt in to tracking or, alternatively, to require payment for access to content if a consumer rejects targeted advertisements.

We believe that this type of choice functionality appropriately balances the privacy interests of consumers with the needs of businesses. It is consistent with consumer preferences as demonstrated by our research described above, and would allow publishers to earn ad revenue or access revenue, which will support high quality website content. The Committee should eschew endorsing





all or nothing tools to enforce Do-Not-Track that would impede consumers' ability to express individualized preferences to specific publishers because publishers are in the best position to determine how to honor consumer choice.

## **ENABLING MARKETPLACE INNOVATION AND MEANINGFUL CHOICE.**

Once a Do-Not-Track mechanism is in place, businesses will be incentivized to provide choice to consumers, and the market will dictate what choices consumers believe are meaningful. Put another way, if consumers require different or more sophisticated choice than a web publisher offers, they will go elsewhere. This will incentivize the market to develop new and better choices to consumers and to engage in best practices. Consequently, the Committee should not be overly prescriptive on what those choices will be – i.e., the Committee does not need to determine which more granular choices, such as category-level (e.g. market segments or product types) or company-level choices, should be available to consumers. The Do-Not-Track mechanism addresses the threshold privacy concern for consumers. The market should be encouraged to innovate and offer more granular choices based on consumer demand. If the approach to choices prescriptive and enables only limited choices, it could have a negative effect on the choices available to consumers and ultimately stifle innovation.

Datran's experience in developing PreferenceCentral is one example of how an appropriate Do-Not-Track mechanism can drive innovation in user privacy. PreferenceCentral offers choice to consumers across brands, rather than ad networks. This has three advantages for the consumer. First, the brand opt-out is honored across all ad networks so that consumers do not have to opt-out of multiple ad networks. Second, by opting out of a brand rather than an ad network, a consumer can continue to get the benefits of tracking – access to



content and personalized ads – while having control over how the collected data is used. Third, consumers find it more meaningful to make a choice over a brand rather than choosing one ad network over another. In this way, PreferenceCentral offers meaningful choice to the consumer because the consumer can continue to receive the advertisements that they want from the brands that they trust. Our research shows that brand control over ads has a meaningful impact on the consumer’s experience and increases their comfort level. As long as the Committee is not overly prescriptive on how choice is to be delivered, consumer friendly services like PreferenceCentral will continue to be developed by many actors in the marketplace.

\*\*\*\*\*

We appreciate the opportunity to be submit these comments and look forward to participating with the Committee.

Steven Vine  
Chief Privacy Officer and Associate General Counsel  
Datran Media

The do-not-track issue is in the middle of two very different, very conflicting interests. While many end users are concerned about being tracked without their knowledge, content providers want (and increasingly depend on) revenues associated with targeted and behavioral advertising. In addition, governments and software vendors have entered the dispute, proposing regulation and implementing HTTP headers to give users the ability to opt out of tracking. However, both solutions by themselves have issues. Trackers could simply ignore do-not-track headers. Regulators could outright ban tracking, which would damage the current model supporting free content through online advertising, or they could set up a Do-Not-Track registry, which would be far more difficult to accomplish with online identities than the Do-Not-Call registry was for static phone numbers, not to mention it would be prone to loopholes. I propose a hybrid solution in which consumers may individually weigh the tradeoffs between privacy and access to free content.

One fundamental component of this hybrid system is the user's right to privacy. That is, the user should have the option to not be tracked without any direct financial cost. I would implement this as the proposed opt-out do-not-track HTTP header. Of course, content providers have the right to be compensated for producing content (if only to cover the costs of producing it), so it would be perfectly valid for a provider to restrict content to those users who do not send the header. This could evolve into a two-tiered system where users either forfeit their privacy in order to access content for free or pay a premium to not be tracked. The user who sends the do-not-track header while still expecting to see content without the premium is bound to be disappointed, but he or she is not necessarily entitled to something for nothing.

However, the user should be allowed some baseline privacy rights even in this "free-with-tracking" tier. Since it would make sense in such a two-tiered system for the default browser settings to not send the do-not-track header (as otherwise the Internet could be a small, closed-off place for the non-tech-savvy), the average user needs some basic protections to prevent being exploited. The authors of such protections could start with banning the obvious — drive-by spyware downloads, external site viewers that bypass browser security, etc. The protections could be amended as needed, to prevent tracking companies from abusing technologies that do not yet exist.

Of course, content providers could just choose to ignore all this without regulatory enforcement. Much of what would go into the baseline privacy rights are already covered through existing laws. However, regulators would also need to require that content providers respect the do-not-track header upfront as well as provide some sort of mode (even if it requires a premium) that respects the header without impacting usability of the website. Such regulations could be part of the many-times-proposed Internet Bill of Rights. Of course, there would have to be provisions defining both when usability becomes negatively impacted and when a premium becomes extortionate, but the latter at the very

least might be solved by the market and/or monopoly law.

While consumer and content provider interests seem to be at odds, it is possible to develop a system that is fair to everyone. Such a system must combine both technological and regulatory solutions to actually be effective. While the content provider and advertising companies are not entitled to exploit end users, end users are not entitled to get something for nothing either. This is just one possible proposal that can balance the interests of both groups.

## About DoubleVerify

DoubleVerify ([www.DoubleVerify.com](http://www.DoubleVerify.com)) is the pioneer of online media verification and compliance. Our mission is to bring trust, accountability, transparency and compliance to online advertising in order to drive the entire industry forward. We are uniquely positioned at the nexus of the digital advertising industry. We work with the largest publishers, advertisers, ad networks, ad exchanges, and demand side platforms to verify the correct placement and display of every single impression.

We are focused on staying one step ahead of non-compliance, inaccuracy and the rogue industry players trying to game the system. We work tirelessly to keep brands safe, advertisers confident and online advertising more accountable and trusted. Our world-class technology and best-in-class client services provide a complete solution for our customer base. DoubleVerify is the approved provider for 6 out of 6 agency holding companies, working with over 150 fortune 500 advertisers and the top 50 networks. We verify and provide compliance on more than 35 billion impressions a month.

## DoubleVerify Memberships, Accreditations and Associations

DoubleVerify is committed to being an industry leader and a central contributor to the online advertising industry. As part of this commitment we have gained the following accreditations and belong to the following associations:

- ✓ DAA Digital Advertising Alliance (DAA) approved provider for OBA compliance
- ✓ DAA advisory board
- ✓ IAB/NAI CLEAR guidelines (main and subcommittees)
- ✓ IAB Network & Exchanges committee
- ✓ IAB & MRC Ad Verification Guidelines Working Group
- ✓ Future of Privacy Forum advisory board
- ✓ Direct Marketing Association (DMA)
- ✓ MRC accreditation (in progress)

## DoubleVerify's Position on Online User Privacy

From the beginning, DoubleVerify has taken a leading position in the online advertising market's Self-Regulatory Program for Online Behavioral Advertising (OBA). We contributed to the development of the Self-Regulatory Principles for OBA, which established the guidelines that the online advertising industry uses to address consumer education about OBA as well as provide consumers control of their privacy with regard to OBA. The principles and program were developed as part of a long-term industry effort to match the intent and direction of consumer privacy protection standards set by the FTC.

We have been an official DAA Approved Provider for OBA compliance since December, 2010. As an approved provider for OBA compliance, we continually demonstrate our commitment to the success of the Self-Regulatory Program for OBA and the online ad industry's dedication to protecting consumer privacy. The Self-Regulatory Program is underway and industry organizations have officially started monitoring and enforcing compliance to it.

DoubleVerify is committed to the success of the Self-Regulatory Program and the market effort to comply with the FTC's direction for consumer privacy and believes that it is vital that any tracking program work in conjunction with the current efforts in regard to OBA compliance.

## **DoubleVerify's Interest in the W3C Workshop on Web Tracking and User Privacy**

The W3C Web Tracking and User Privacy workshop addresses many of the same concerns that DoubleVerify is focused on resolving; therefore, we consider it essential to participate so that together we can develop methods and guidelines that support and enhance current efforts made by the industry to meet the best practices called for by the FTC.

DoubleVerify is looking forward to working with the W3C Web Tracking and User Privacy workshop. We believe collaboration between the concurrent efforts for market compliance that is being pushed by the DAA and the W3C will ensure that both programs work together to enrich the market and enhance our overall capabilities.

DoubleVerify's rich experience with online media—specifically online verification, compliance and privacy—makes us a perfect partner for sharing our input on the current methods used for OBA compliance and ad verification. Our experience allows us to provide insight into discussing implementations and solutions for a 'Do Not Track' mechanism.

# Trackers Don't Track People, People Track People or What We Really Mean When We Say "Do Not Track"

A Position Paper from W3C Workshop on Web Tracking and User Privacy  
Andy Kahl and Colin O'Malley - Evidon, Inc.

In 2003, Congress passed and then-presented George W. Bush signed into law the Do Not Call Implementation Act, which mandated that the Federal Trade Commission create and maintain the National Do Not Call Registry. It was an extremely popular measure, and why wouldn't it be? Unsolicited marketing calls are categorically invasive. Adding a number to the registry means that, with few exceptions, unsolicited calls to that number are forbidden. It is a simple, analog decision and the legislation that allowed it was both timely and effective. At a glance, an initiative that would allow users to opt-out of online data collection seems very similar, and so proposals to limit or disallow these practices have earned the collective nickname "Do Not Track". There is a natural tendency to also think of the solution in satisfyingly similar terms. Many of the proposed Do Not Track solutions, therefore, focus exclusively on the blocking or opting-out of data collection. While meaningful options of this nature are important components to an effective solution, they are not a solution in its totality. A privacy-conscious, ad-supported Internet requires transactional transparency, relevant information, and meaningful choices for the end-user.

Transparency is key on several levels - not the least of which to counteract the idea of creepiness that is often repeated in criticism of online data collection. In real-world transactions, shop owners learn your name, buying preferences, and other relevant details in order to customize your experience. Far from uncomfortable or "creepy", this kind of service is heralded as attentive and valuable. If those same shopkeepers quietly looked at your other shopping bags to guess at your purchase history, shared what you bought with other stores, and used your driver's license information to look up details about your family; you would quickly move from satisfied to disturbed. Anytime data that was not explicitly provided is explicitly used, there is a reflexive notion of privacy violation. The use of the data is not as problematic as the opaque nature of its collection. Transparent collection helps build a sense of trust and avoids giving users the creeps when that data is subsequently used. This is a general and systemic policy-based move, but one that is nonetheless critical.

The idea of transactional transparency is much more specific. This means notifying the user that data use or collection is part of a given user action. These actions could be page loads, ad delivery, widget execution, etc. This is a particularly important and inescapable feature of a robust and functional ecosystem, as users cannot build trust relationships with companies in the industry if they are not aware of when and how their data is used. It is noteworthy that many of the opt-out mechanisms discussed as part of Do Not Track proposals fall far short of the goal of transactional transparency. Technology that blocks as part of a list, obfuscated browser options, or by setting and maintaining opt-out cookies may offer the user a sense of control, but collectively lack a persistent indicator that the user has decisions to make. This

could easily result in a false sense of privacy for end users, as data use and collection will continue in cases that the user isn't protected by an invisible list or opt-out cookie. That false sense of security is exacerbated when a one-click, Do Not Track mechanism exempts large categories of commercial entities, as many of the discussed proposals have. These exemptions are often warranted and reasonable; but without transparent and real-time notification, it is easy to envision a consumer believing that they have opted-out of whole types of tracking that are actually excluded from the Do Not Track feature.

It is, in essence, unreasonable to assume that an effective system can be created that does not include real-time, transparent user notification. To guarantee that users are well informed and are making active decisions about their data, users should be clearly notified every time data collection or use is attempted, even if they have previously opted-out. A solution without notification is particularly risky for publishers. An ideal system is one in which users decide if content on a given page valuable enough to allow for some data collection. Publishers, in turn, closely manage their partnerships with advertising companies to ensure that user data is only being used for this purpose. Without a system that standardizes transparent notification, users have no way to judge one site's data collection practices against another. Data collection becomes taboo instead of currency, and fundamental changes are required in the way online content is subsidized.

Transparent notification is only valuable when attached to relevant information. Simply displaying an icon that tells users "You've Been Tracked!" is not a legitimate aide to user privacy. They must have the opportunity to make informed choices. The informed nature of that choice is critical, which makes relevant information a core component of an effective solution. Data collection varies widely in both policy and practice. Companies differ on what data is collected, how that data is used, whether it is shared, how long it is stored, and to what extent the user can alter the data about themselves. The technology for data collection also varies widely, from server-side storage of elements like search strings to cookie-based session storage of a user's reaction to an advertisement, and many implementations in between. Some companies offer robust preference management where the user can shape the data collected. Even opt-out choices are variant, as an opt-out to one company doesn't mean the same as it does to another. Users need a real-time understanding of the companies involved in data collection on a given site, their policies, and then the choice to opt-out (coupled with an explanation of what that means). From a policy perspective, this information should be easy to understand and relevant to the actions taking place. It is critical that we learn lessons from previous failures in user notification like financial disclosures in user agreements from bank and credit card companies. Large dumps of standardized information anytime a user is notified of a data collection action undermines the value of transparency and does not allow for an informed decision by user. It's established that transparency is necessary in principle, and this transparency must be continued in practice through the provision of relevant, digestible, meaningful data.

A notified, informed user should then be allowed to make a meaningful choice. The data control offered by this choice should be clear, and the choice should be as close to permanent



as possible. The core issue here is one of policy, not technology. It is certainly possible to release technology that uniformly blocks the common tools that data collection companies use to operate, but technological hurdles are easily circumvented, and cannot be regarded as a solution on their own. Data collection companies must adopt policies that result in a contractual understanding between their operation and consumers. Consumers must be offered a decision - and the must be given the opportunity to weigh the benefits of both sides of that decision. A permanently affixed "Not Me" sign is not a representation of an engaged, meaningful choice; and neither is a convoluted and token opt-out policy that offers consumers very little actual control. It is not outlandish to assume that this trade-off can be expressed in a way that allows a consumer to understand risk versus value, and subsequently make a material choice based on that understanding.

Collectively, these efforts can create an online ecosystem that is simultaneously advertising supported and privacy sensitive. Further, it supports a system of buyers and sellers, not creepy conspirators and their hapless victims. Informed, active consumers who understand the value of their data can leverage that currency in the same way they leverage their offline currency. The era of the friendly online shopkeeper is within our reach.

Author contact information:

Andy Kahl - [andy@evidon.com](mailto:andy@evidon.com)

Colin O'Malley - [colin@evidon.com](mailto:colin@evidon.com)

[www.evidon.com](http://www.evidon.com)

## Future of Privacy Forum W3C Do Not Track Position Paper – Request to Participate

Jules Polonetsky is Director of the Future of Privacy Forum, an industry supported Washington based think tank which includes an advisory board of advocates, academics, data privacy regulators and Chief Privacy Officers. Jules was previously the Chief Privacy of AOL and of DoubleClick, as well as the Consumer Affairs Commissioner of New York City.

Shaun Dakin is a Fellow at the Future of Privacy Forum, where he focuses on privacy issues related to applications. Shaun is also the Founder of PrivacyCamp and the National Political Do Not Call list.

Since its founding in 2008, the Future of Privacy Forum has played a role in helping advance consumer friendly and business practical online privacy practices. We designed and consumer tested an icon for companies to use to indicate that an ad is behaviorally targeted. We have coordinated a group of companies focused on improving the current cookie based opt-out process, and have held several public and non-public programs to advance conversation of the Do Not Track header. We have generated data flow charts to explain how information is used by the range of participants in the online marketing ecosystem and act as an expert resource for media, policymakers and leading companies. Jules was a member of the W3C working group involved in the development of the P3P standard and was a founder of the Network Advertising Initiative, the group of ad networks that set in place the original self-regulatory program requiring ad networks to provide cookie based opt-outs for behavioral advertising.

The Future of Privacy Forum has argued that the Do Not Track header can play a key policy based role in advising companies that a particular user doesn't want to have their data used for more robust personalization or marketing purposes. We think the current opt-out process relying on cookies is faulty and unreliable. In the mobile and app ecosystems, where a range of methods are used for tracking and where cookies aren't always available (or are limited for first party use), the cookie related opt-out for behavioral ads is of very limited value. We think that it will be useful for a Do Not Track header to be cognizant of the relevant self-regulatory programs to be able to leverage industry acceptance and oversight. But we also think that there may be areas where industry programs have been unclear about scope and efforts to use a header may provide opportunities for further progress.

We also think that Do Not Track headers should not be viewed as a technical privacy solution, replacing enhanced cookie controls or other private browsing features that will continue to develop, nor as a comprehensive solution for all online privacy issues. Rather, it should recognize that the same types of data maybe collected by companies for purposes with a range of privacy impact, whether for very limited analytics or very robust targeting and sharing. Technical controls that seek to limit data will either over block or underblock. A header is best

viewed as a policy based solution that can be described in a very simple and clear manner to mass audience users with an effect they can understand. For example – don't let other companies tailor ads for me based on my visit to a web site. Complexity or great detail and nuance in this area may complicate consumer understanding and diminish the value of the use of the header.

Given our position as a convenor, and an entity that takes input from a range of industry, advocacy and academic actors, the above position is not a formal position of the Forum. Our main goal is to advance privacy practices in a manner that provides additional control and transparency for consumers, while supporting responsible uses of data. Jules would be pleased to participate at the program in any way useful, but may be particularly useful as an active moderator, a role he very often plays.

Contact Info:

Jules Polonetsky, Director

Shaun Dakin, Fellow

Future of Privacy Forum

919 18<sup>th</sup> Street, suite 925

Washington DC 20006

202 713-9466

[julespol@futureofprivacy.org](mailto:julespol@futureofprivacy.org)

[shaun@futureofprivacy.org](mailto:shaun@futureofprivacy.org)

## A Social Network Users' Bill of Rights: "You" Must Decide

Christina M. Gagnier, Esq.  
Gagnier Margossian LLP  
[gagnier@gamallp.com](mailto:gagnier@gamallp.com)

A Social Network Users' Bill of Rights provides a framework for town hall style engagement in the complex online privacy policymaking process while arriving at a body of generally accepted principles that can guide government regulatory efforts. The importance of such dialogue is in addressing the normative aspects to privacy law and user rights, whether reinforcing the tort protections currently available for individual privacy here in the United States or working towards a set of "principles" that recognizes the international impacts of web platforms. A norm-driven endeavor can support entrepreneurs to employ architectural solutions to privacy, creating a complimentary regulatory framework and marketplace that rewards "privacy by design" while promoting innovation.

### **I. Introduction**

The convergence of "reality media," social networks and instant publication have led to the misconception that privacy is dead; rather, we remain in our own societal beta test of the global power of the social net. For digital natives and the rest of us, valuing privacy has always required some sort of a contextual element, a lesson, a moment or an incident that causes one to value privacy. Privacy is subject to a trigger effect.

The trigger for many came with the Facebook and Google privacy controversies of 2010. The technological changes by these companies that undermined user privacy led various stakeholder groups to come together at the 2010 Computers, Freedom and Privacy Conference and create a draft "Social Network Users' Bill of Rights." The "Bill of Rights" (#snubor), in its current draft form, includes the following rights:

*"We the users expect social network sites to provide us the following rights in their Terms of Service, Privacy Policies and implementations of their system:*

- 1. Honesty: Honor your privacy policy and terms of service.*
- 2. Clarity: Make sure that policies, terms of service, and settings are easy to find and understand.*
- 3. Freedom of speech: Do not delete or modify my data without a clear policy and justification.*
- 4. Empowerment: Support assistive technologies and universal accessibility.*
- 5. Self-protection: Support privacy-enhancing technologies.*
- 6. Data minimization: Minimize the information I am required to provide and share with others.*
- 7. Control: Let me control my data, and don't facilitate sharing it unless I agree first.*

8. *Predictability: Obtain my prior consent before significantly changing who can see my data.*

9. *Data portability: Make it easy for me to obtain a copy of my data.*

10. *Protection: Treat my data as securely as your own confidential data unless I choose to share it, and notify me if it is compromised.*

11. *Right to know: Show me how you are using my data and allow me to see who and what has access to it.*

12. *Right to self-define: Let me create more than one identity and use pseudonyms. Do not link them without my permission.*

13. *Right to appeal: Allow me to appeal punitive action.*

14. *Right to withdraw: Allow me to delete my account, and remove my data.”*

At the Southwest by Southwest Interactive festival in March 2011, accompanying two of the “Bill of Rights” authors, Lisa Borodkin and Jack Lerner, the “Bill of Rights” was presented in an open discussion. The points outlined within this paper address the commentary provided, and the questions asked both on and offline about the contents of the draft itself. This content is fundamental to the debate surrounding user privacy. Despite treatment by the Federal Trade Commission and the Department of Commerce in their policy papers and emerging legislative solutions, most recently, the McCain-Kerry draft online privacy bill, an effort to actively solicit public input as to desired privacy and associated user rights is lacking in this continuing evolution of consumer rights online.

## **II. The lack of an individual private right of action within circulated legislative draft bills further disables individuals to rely on the tort protections available for the protection of individual privacy in the absence of a mechanism to evaluate user norms.**

Our tort privacy framework in the United States is largely shaped by cultural norms and practices. The lack of a private right of action within the circulated draft provisions of the McCain-Kerry online privacy legislation, and the absence of discussion of these protections in the policymaking process, merits user-driven efforts. Legislative attempts aim to address the obvious violations of user trust and data mismanagement of platform providers, but fail to recognize the underlying priorities users may have. The prioritization by a broad and diverse user base of the rights they deem necessary, and in some cases, fundamental to their use of social networks, will serve to educate the legislative process and create a normative framework necessary to perhaps revive the tort protection currently available for individual privacy. A legislative fix is not necessarily a normative one, and does not serve the enhanced need for users to access these tort protections in the digital age.

What is a “reasonable expectation of privacy” in our digital age? User demands at times seem incompatible: users openly share photos of intimate situations while demanding privacy of certain personally identifiable information. In the battle to coalesce the tension between the virtual and visceral world, the concepts who is a “public figure,” what is of “legitimate concern to the public,” and generally determining the difference between private and public space are in flux. These are not merely abstract ideas: they are the elements of existing tort actions relied upon by individuals to protect their privacy rights. It is imperative to learn from users and develop norms around these concepts, as their legal impact comes from a judicial interpretation of shifting societal mores.

Aside from merely protecting these individual rights in terms of our existing privacy tort regime, the “rights” conversation is not limited to user rights vis-à-vis platforms: there is a need to address the interactions between users as well. An intimate understanding of the desired rights in these situations can only come from users themselves; a top-down legislative fix will not serve to resolve these points of contention when the technology shifts so rapidly and some challenges may be platform-specific. Each platform has its own unwritten “code,” a set of norms and language standards that only apply in that context. Understanding the necessary data limitations and privacy expectations that should be placed on these platforms must be generated from the users themselves based on collective experience.

**III. A user-generated effort allows for the integration of norms beyond those developed in the United States, supporting a movement towards universal human rights principles online.**

Organizing around a “Social Network Users’ Bill of Rights” accomplishes normative goals that are essential to our existing privacy regime in the United States. Yet, this conversation has international implications: it is platforms provided by companies based in the United States that are defining rights for a growing base of international users.

A gaping void in international human rights law, the lack of treaty law and customary international law concerning human rights in the online world, leaves those who are denied basic freedoms very little to turn to. While principles protecting free expression are codified in both the *Universal Declaration of Human Rights* and the *International Covenant on Civil and Political Rights*, the borderless nature of the Internet, and the daily global interaction between Internet users, transforms the domestic online privacy conversation in the United States into an international one.

The powerful utility of social networks in facilitating mass communication during political revolution could not have been envisioned less than two years ago, beginning with Iran, a situation where content spread a message virally across the globe in a handful of hours. Hashtags like #jan25 support the idea of a continued “rights” conversation, as our society finds itself in the early stages of understanding what rights and values are important in an international context, with the consideration that many of these platforms are provided by companies in the United States.

Platform agnosticism is one such issue that has emerged, a proposed “right” that would prevent companies from taking sides in international conflicts by filtering or removing content, inspired by companies like Twitter who have taken the lead by demonstrating such a commitment. In the larger international law conversation, the treatment of these networks as non-state actors, with a public function or nexus in times of social and political turmoil, has yet to be explored.

**IV. While a subset of bad actors has necessitated the need for government intervention in online privacy, a “Bill of Rights” provides an opportunity for entrepreneur-supported efforts, allowing for the creation of voluntary mechanisms leading to a competitive privacy marketplace with a complimentary regulatory framework.**

Media has embraced the narrative of a necessarily adversarial relationship between user and platform. This context has caused many emerging startups to perceive the attempts at regulation as anti-entrepreneur: additional rules and regulations will hinder growth, whether or not one finds merit in the “technology bubble” rumors around the startup industry. Regulation may be necessary for the “800-pound gorilla” bad actor, but a policymaking effort absent startups, specifically in the burgeoning data industry, may stymie essential economic growth.

No doubt governments, at the state and federal level, find themselves in a conundrum with user rights: a “wait and see” approach, hoping that the bad actor will change its policies, brings with it few guarantees. Additionally, positive incremental changes to the privacy practices of platform operators are not indicia of a corporate commitment to respecting user rights. When Facebook announced in October 2010 that it was now allowing users to download their data from their Facebook account, putting users in control of where their data is and in the driver's seat for where they take their data next, these actions seemed to comport with rights that were included in the draft Social Network Users’ Bill of Rights, specifically, Article 7, "Data Control," and Article 14, "Right to Withdraw.” Despite these changes, the six-month time period since these the launch of these features has been riddled with announcements by Facebook of practices exposing its users to an increased amount of data misuse.

Competition with privacy is a tricky concept; it cannot replace the need for regulation, but regulation and competition can surprisingly be complimentary. Some privacy adherents believe true preservation of privacy rights can only be achieved through “code,” through site architecture that is designed with privacy in mind. The market may be the appropriate place to provide these incentives to startup companies while a complimentary regulatory framework, cognizant of the challenges faced by entrepreneurs, can police the companies whose actions will have the largest net impact on the user community.

## **V. Conclusion**

The corporate-created legal regimes that users are currently beholden to have their failures, but the mechanisms to address these failures are still in the early stages of development. The Social Network Users’ Bill of Rights states in its preamble, “We the users...” In order to create a true solution in terms of protecting user rights, user stakeholders must be involved.

# Transparency and Choice: Protecting Consumer Privacy in an Online World

Alma Whitten<sup>a</sup> Sean Harvey<sup>b</sup> Ian Fette<sup>c</sup> Betsy Masiello<sup>c</sup> Jochen Eisinger<sup>d</sup> Jane Horvath<sup>e</sup>  
{alma,sharvey,ifette,betsym,eisinger,janehorvath}@google.com

<sup>a</sup> Google Inc., Belgrave House, 76 Buckingham Palace Road, London SW1W 9TQ, UK

<sup>b</sup> Google Inc., 76 Ninth Avenue, New York, NY 10011, USA

<sup>c</sup> Google Inc., 1600 Amphitheatre Parkway, Mountain View, CA 94043, USA

<sup>d</sup> Google Germany GmbH, Dienerstr. 12, D-80331 München, Germany

<sup>e</sup> Google Inc., 1101 New York Avenue, N.W., Second Floor, Washington, DC 20005, USA

## Abstract

There have been concerns raised recently about online tracking. There are a variety of mechanisms by which data is collected online, and for which it is used, and it is unclear which of these are intended to be addressed by “Do Not Track” mechanisms. Tracking is often data collection that helps ensure the security and integrity of data, determines relevancy of served content and also helps create innovation opportunities. This value ought to be central in any “Do Not Track” discussions.

## 1. Introduction

The idea of a “Do Not Track” mechanism has inspired debate among those concerned about online tracking. Several mechanisms and solutions exist or have been proposed to provide users with choice and control over the profiling they experience online. Each of these has limitations and consequences—none appears to be a panacea for concerns about tracking. Perhaps most significantly, there is a wide range of definitions of “tracking” and thus no uniform problem to solve for. If implemented carelessly, solutions to prevent tracking could have the unintended consequence of diminishing the online experience for users and stifling the growth of online publishing without meaningfully improving user privacy.

## 2. Tracking

There are many types of data collection that occur when a user browses the Web, and these occur for many different reasons. In discussions about “Do Not Track” it is important to be concrete about what is meant by “track.”

Mechanisms of data collection that create streams of information about a particular user or browser include HTML cookies, javascript, authentication, or advanced types of “fingerprinting.” Data is



collected by first-party publishers to serve personalized content to users. It can also be collected by third-party content providers for the same reason. Data may also be used to monetize content, either by serving contextually targeted advertisements to a user, or by inferring interests that a particular browser or user is likely to have and serving ads targeted to those interests. Importantly, the data collected may be used not only to personalize content and advertisements, but also to protect and secure services from fraud and abuse.

### **3. Existing approaches**

There are a variety of existing approaches to preventing tracking. The longest-standing approach relies on cookie settings in the browser: users of most major browsers can choose to block all cookies from being set, block third party cookies from being set, or in some cases block cookies from specific domains from being set. More recently a number of approaches have sought to build on the cookie infrastructure but enable users a more global option for preventing ad targeting.

A second variety of approach is browser extensions and features that block network requests altogether or block the display of network content. This network-level blocking is effectively based on a list of domains to which network requests cannot be sent if an extension or feature is turned on, or from which content cannot be viewed.

A third and newly proposed approach is an HTTP header. The idea is for users to signal their preference to not be tracked universally to all websites they visit.

### **3. Analysis**

What is sometimes referred to as tracking is often data collection that helps ensure the security and integrity of data, determines relevancy of served content and also helps create innovation opportunities. It is important not to let a single negatively-loaded term obscure the fact that data collection is the source for the creation of value as well as the legitimate concerns of different parties.

A common assertion made in discussions about tracking is that average users do not want to be tracked and do not understand the tracking ecosystem. As we set out goals for the workshop, one question to ask is whether one goal should be helping users understand the data collection that occurs online as well as its risks and benefits.

One observation of the range of solutions described above is that most, though not all, focus on providing the user a simple decision interface for an inherently complex ecosystem: turn on a header or don't, block third party cookies or don't, install a content blocking feature or don't. However, some solutions have features that focus on enhancing transparency to the end user. We should analyze the impact of improved transparency on the user's online experience, as well as their understanding of the ecosystem, and whether improved transparency influences the decisions they make about various

forms of tracking.

In addition to focusing on the user experience, there needs to be a focus on user value. An important observation is that the tracking used to monetize online services may be invisible to the user, and yet provide the user immense value. Advertising may be less annoying or intrusive if it is useful and relevant to the user. As a simple example, data collection enables advertisers to do frequency capping, which ensures that the same ad is not shown repeatedly to a given browser or user. Tracking allows advertising companies to monitor for fraudulent services or deceptive ads, further ensuring that irrelevant messages or offers are not intruding a user's experience. To ask a user to make a decision about tracking without incorporating both sides of the equation—the value they get from advertising-supported content as well as their concerns about tracking—would put at risk aspects of the online experience that millions of users have come to expect and value.

#### **4. Conclusions**

To address the concerns of users and give them effective tools to improve their online privacy, it is important to (a) be transparent about what data is collected and how it is used, and (b) offer users meaningful choices that are understandable to both users and sites being notified of these choices, without mysterious or unintended side effects. A commonly agreed-upon definition of tracking would be an important step forward. Tracking ought to consider the connection of information about online behavior to the offline world, as well as the rise of cloud-based computing and the ever growing mobile market. Most importantly, solutions should put the user first: what should the user understand about tracking if they are to make an informed decision, what expectations do they have about their online experience that they may unknowingly compromise without that understanding, and what is the value that users derive from tracking?

# Do-not-track as a driver for transparency of social networking advertisement practices?

**Jens Grossklags**

College of Information Sciences and Technology  
The Pennsylvania State University  
University Park, PA 16802

A number of different approaches have been utilized to monetize social network data and human capital. Least controversial are fan pages on social networking sites created by companies. Those efforts stimulate brand awareness, loyalty and foster a direct communication channel between a company and its potential customers. In a study of online retailers, about one-third self-reported that they maintained a Facebook page, 27% had a MySpace site and 26% created a presence on YouTube (Internet Retailer & Vovici, 2008).

However, the utilization of social network data for targeted advertisements is considered highly contentious. In a recent survey study, 66% of the surveyed adult Americans and 55% of the 18-24 year-old young adults prefer marketers to abstain from such efforts (Turow *et al.*, 2009). But behavioral and targeted advertisement is effective. 63% of the senior marketing executives report that it yields the greatest return on investment. At the same time, at least some companies are scaling back investments into related technologies as a result of consumers' privacy concerns (Ponemon Institute, 2010).

This reluctance can be explained given the state of the art of the marketing research literature: It is still hard to predict when consumers will *welcome*, *acquiesce*, or vigorously *protest* against new practices. Customers, who are burned once, may be twice as shy down the road to interact with marketers (Good *et al.*, 2005). Google and Facebook have weathered the storms that resulted from the release of Buzz and Beacon, respectively, but smaller content providers may not be so fortunate.

Indicators of consumer response may be delayed or subject to factors that are typically not accounted for in advertisement effectiveness studies. In our previous experimental work, we showed that consumers may regret their own decisions and feel betrayed even though they initially seemed to allow certain marketing practices and privacy invasions (Good *et al.*, 2007). Similarly, individuals' stated preferences may significantly differ from their eventual behaviors in marketing contexts (Spiekermann *et al.*, 2001). Related research contributes other puzzling revelations. For example, advertisements that are relevant to the website content *or* are obtrusive increase willingness to purchase. But a combination of these two factors is counterproductive (Goldfarb and Tucker, forthcoming). In another study, pop-ups were shown to increase brand awareness, but also

to reduce reservation prices (Acquisti and Spiekermann, forthcoming). These researcher groups speculate that certain practices may trigger consumers' feelings of *manipulation and deception* (Boush *et al.*, 2009).

Further, research fails to account for current practices utilizing social networking data in static and dynamic ways. In the former case, such data is frequently used as endorsements in advertisements on unrelated sites (including offline marketing efforts). New campaigns (including HP's) often include comments from Twitter and Facebook in rich banner ads (Dilworth, 2010). In the latter case, Facebook's new social plugins push user data to a wide variety of websites to offer *instant personalization* (Gannes, 2010).

It is reasonable to assume that consumers are neither fully aware of different advertisement trends nor completely understand the different means and ways of how their data is collected, shared and eventually utilized (Stein, 2011). One potential response is to aim for a higher degree of *transparency* with respect to advertisement practices involving social data.

The proposed Do Not Track Me Online Act of 2011 (H.R. 654) would not only give consumers a measure of control over data treatment, but also calls for entities that are affected by the new law to disclose their practices for collection and sharing, including the identities of data exchange affiliates. And, in anticipation of regulatory changes at least one major advertisement intermediary has started a pilot project to improve transparency and relevance (Wilson, 2011).

It is less obvious whether these trends will lead to more meaningful options for consumers and choices by consumers.

First, in the short term, the plethora of potential do-not-track implementations is likely not going to converge on a simple and effective market standard. Yu's (2010) discussion of design choices clearly highlights the problems ahead. On the one hand, the technical details of implementations can severely thwart the real-life impact of do-not-track. For example, different ways to aggregate externally provided blacklists for overly aggressive marketers in the browser can appear unintuitive for the consumer and even technologists (Clarke, 2011). On the other hand, conceptual problems are in need to be tackled by researchers. In particular, the trade-off between simplicity (e.g., a binary on/off choice) and fine-grained preference management is challenging from a variety of perspectives as evidenced by the discussions around privacy management, e.g., in the context of the Platform for Privacy Preferences (Cranor *et al.*, 2002).

Second, given the concentration in the advertisement industry one has to carefully observe whether the given data management options translate into meaningful consumer choices. The idea of do-not-track is inspired by regulatory efforts that are considered highly successful from a consumer protection perspective such as the do-not-call registry (Varian *et al.*, 2005). But the achievements of the do-not-call registry do not only rely on its simplicity (including the semi-permanent nature of telephone numbers) but also on the dynamics of the interactions that are concerned. Specifically, it mainly addresses unsolicited calls within the confines of the privacy of the home while consumers are engaged in their *private unrelated affairs*. In contrast, do-not-track is closely tied to interactions that are initiated by the consumer and deeply embedded in popular activities such as partaking in a social network, shopping on an ecommerce site, or information

gathering on news outlets. Companies offering these requested services have a reasonable expectation to benefit from their offerings. And consumers may feel constrained in their effective choices when they are related to services with strong network effects or market dominance. Further, these impediments will likely influence consumer behavior also on sites that do not fit these criteria.

Do-not-track will lead to more transparency in the advertisement industry, whether through regulatory actions or industry-guided efforts. However, research needs to be undertaken to understand whether this trend will help to overcome consumer privacy hurdles.

## References

A. Acquiti and J. Grossklags, Privacy and rationality in decision making. *IEEE Security & Privacy*, 3(1):24–30, 2005.

A. Acquiti and S. Spiekermann, Do Pop-ups Pay Off? Economic effects of attention-consuming advertising, *Journal of Interactive Marketing*, forthcoming.

D. Boush, M. Friestad, and P. Wright (2009) *Deception in the marketplace: The psychology of deceptive persuasion and consumer self protection*, Routledge, New York, NY.

G. Clarke, Microsoft: IE9's web privacy hole? A feature, not a bug - When do-not-track lists clash, *The Register*, March 18, 2011.

[http://www.theregister.co.uk/2011/03/18/microsoft\\_ie9\\_tpl\\_site\\_blocker/](http://www.theregister.co.uk/2011/03/18/microsoft_ie9_tpl_site_blocker/)

L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle. The Platform for Privacy Preferences 1.0 (P3P1.0) Specification, April 2002.

<http://www.w3.org/TR/P3P/>

L. Gannes, Facebook: The Entire Web Will Be Social, GigaOM, 2010.

A. Goldfarb and C. Tucker, Online Display Advertising: Targeting and Obtrusiveness, *Marketing Science*, in press.

N. Good, R. Dhamija, J. Grossklags, S. Aronovitz, D. Thaw, D. Mulligan and J. Konstan, Stopping Spyware at the Gate: A User Study of Privacy, Notice and Spyware, *Proceedings of the Symposium On Usable Privacy and Security (SOUPS 2005)*, Pittsburgh, PA, July 6-8, 2005, pp. 43-52.

N. Good, J. Grossklags, D. Mulligan, and J. Konstan, Noticing Notice: A large-scale experiment on the timing of software license agreements, *Proceedings of the ACM Conference on Human Factors in Computing Systems (CHI'07)*, San Jose, CA, April 28 - May 3, 2007, pp. 607-616.

Internet Retailer & Vovici, 2008. Emerging Technologies. Market study.

Ponemon Institute, 2010. Fear and Loathing in Online Advertising, Research Report.

S. Spiekermann, J. Grossklags, and B. Berendt (2001) E-privacy in 2nd generation e-commerce: Privacy preferences versus actual behavior, *Proceedings of the Third ACM Conference on Electronic Commerce (ACM EC'01)*, pp. 38-47.

J. Stein, Data Mining: How companies now know everything about you, *Time*, March 10, 2011.  
<http://www.time.com/time/printout/0,8816,2058114,00.html>

J. Turow, J. King, C. Hoofnagle, A. Bleakley and M. Hennessy, Americans Reject Tailored Advertising and Three Activities that Enable It, University of Pennsylvania and University of California, Berkeley, report, 2009.

H. Varian, G. Woroch and F. Wallenberg, The demographics of the do not call list. *IEEE Security and Privacy*, 3(1):34-39, 2005.

D. Wilson, Yahoo launches a transparent advertising scheme: Claims it's a better consumer experience, *The Inquirer*, March 18, 2011.  
<http://www.theinquirer.net/inquirer/news/2035321/yahoo-launches-transparent-advertising>

H. Yu, Do Not Track: Not as Simple as it Sounds, *Freedom to Tinker (Blog)*, Princeton University, August 10, 2010.  
<http://www.freedom-to-tinker.com/blog/harlanyu/do-not-track-not-simple-it-sounds>

# **An Evaluation of Self-Regulation of Consumer Tracking and Profiling: Deficiencies and Recommendations for Improvement**

Submission to W3C Workshop on Web Tracking and User Privacy

28/29 April 2011, Princeton, NJ, USA

Keyna Chow, Nicholas Petersen & Chris Jay Hoofnagle

Samuelson Law, Technology & Public Policy Clinic

March 24, 2011

## **I. Introduction**

This position paper is written on behalf of The World Privacy Forum (“WPF”) by members of the Samuelson Law, Technology & Public Policy Clinic at University of California, Berkeley School of Law. The WPF is a nonprofit, non-partisan 501(c)(3) public interest research group that works both nationally and internationally. The organization is focused on conducting in-depth research, analysis, and consumer education in the area of privacy.

Our project<sup>1</sup> evaluates self-regulatory principles<sup>2</sup> for online advertising, highlights their critical failings, and recommends alterations to the codes that better suit consumer privacy interests. Understanding these codes is critical for technologists, because technological approaches to web tracking and user privacy will be complemented by a mixture of self-regulatory norms and enforcement by the Federal Trade Commission (“FTC”).

At the April workshop, (1) we will explain how the policy debate on industry self-regulatory programs for online advertising is relevant to technologists, and (2) we will explore with the workshop participants whether flaws in self-regulation are purely policy issues or whether there are technical solutions to these challenges. Our presentation will emphasize:

- Political framing of the phrase “Online Behavioral Advertising” – does the phrase “Online Behavioral Advertising” (“OBA”) adequately address privacy concerns? Does that phrase accurately describe what is happening in a technical sense?
- Flaws and loopholes in the Network Advertising Initiative (NAI) and Digital Advertising Alliance (DAA) programs – can technology fill their policy gaps or do we need to strengthen the policy itself?

---

<sup>1</sup> On behalf of the World Privacy Forum, our project will produce a white paper evaluating the various self-regulatory programs for online behavioral advertising in May 2011. In addition, we submitted a comment to the Senate Commerce Committee discussing the key changes with the new NAI (*The NAI Then And Now: What Has Changed In Advertising Self-Regulation*, February 16, 2011) and a comment to European Advertising Standards Alliance (EASA) on its Best Practice Recommendations on Online Behavioral Advertising (*Comments on EASA Best Practice Recommendation on Online Behavioural Advertising*, February 25, 2011).

<sup>2</sup> We will evaluate, at a minimum, the NAI and the DAA principles in our project.

## II. Background on Self-Regulation of Online Advertising

Our prognosis for the self-regulatory endeavor is bleak. Even at the most surface level, self-regulatory proposals fail to fully embrace the consumer privacy interests at stake.<sup>3</sup> For example, the DAA's Self-Regulatory Principles for Online Behavioral Advertising does not even invoke "privacy" as a policy goal. It only refers to privacy descriptively (e.g. to identify "privacy policies" and the like), and does not recognize consumer privacy as a legitimate interest until page 35, where Internet service providers engage in Deep Packet Inspection.<sup>4</sup>

Without privacy as a policy goal, substantive provisions fail to address the most pressing issues at hand. Take, for example, Network Advertising Initiative's ("NAI") definition of the practice of online behavioral advertising: "OBA means any process used whereby data are collected across multiple web domains owned or operated by different entities to categorize likely consumer interest segments *for use in advertising online.*"<sup>5</sup> This means NAI participants could be compliant, collect consumer data, and use that data for other purposes, so long as it is not for advertising online.

More broadly, tracking and profiling implicate consumer privacy whether or not data are used for advertising online. For the consumer, this means that opt-out abilities are largely illusory, because it only restrains *use* (not collection) of information for advertising purposes. The Principles sabotage the objectives that motivated intervention to begin with, dodge consumer concerns, and they thereby undermine the credibility of the self-regulation program. This and other self-regulatory codes stand upon a flawed foundation, and their regulatory codes reflect those foundational flaws.

## III. Analysis

A credible self-regulatory scheme should meet certain minimum standards of independence, accountability, and structural features to maintain legitimacy. Our project explores what makes a good self-regulatory program and how well the various codes meet those expectations. We would like to present this research at the April workshop. Our preliminary research has shown that the self-regulatory codes submitted by the advertising industry are flawed and have ample room for improvement with regard to independence, accountability, and other self-regulation best practices.

---

<sup>3</sup> National Consumer Council, *Models of self-regulation: An overview of models in business and the professions* 23 (November 2000), available at [http://www.talkingcure.co.uk/articles/ncc\\_models\\_self\\_regulation.pdf](http://www.talkingcure.co.uk/articles/ncc_models_self_regulation.pdf) ("The objectives must be rooted in the reasons for intervention"; "The scheme must be based on clear and intelligible statements of principle and measurable standards – usually in a Code – which address real consumer concerns." (emphasis added)).

<sup>4</sup> Digital Advertising Alliance, *Self-Regulatory Principles for Online Behavioral Advertising*, July 2009, <http://www.aboutads.info/resource/download/seven-principles-07-01-09.pdf>.

<sup>5</sup> 2008 NAI Principles: *The Network Advertising Initiative's Self-Regulatory Code of Conduct*, pg 4, emphasis added, [http://www.networkadvertising.org/networks/2008\\_NAI\\_Principles\\_final\\_for\\_Website.pdf](http://www.networkadvertising.org/networks/2008_NAI_Principles_final_for_Website.pdf)



Independent scrutiny of advertisers is necessary to ensure consumer privacy is adequately protected.<sup>6</sup> The proposed self-regulatory codes are not designed to protect privacy interests in a way that counteracts the financial incentives of online behavioral advertising. They lack the ability to critique their members unencumbered by their reliance on those members' financial support. Furthermore, the program does not command a sufficient share of the advertising market to make membership a prerequisite to doing business. The self-regulatory codes must take steps to expand membership so that they can overcome their financial reliance on the industry.

Embedded in the industry self-regulatory programs is a binary approach to privacy that can leave the consumer completely unprotected based upon choice or consent. For instance, the programs call for explicit consent to the adoption of technologies such as browser toolbars. Nothing in the programs calls for these technologies to be cabined through privacy-by-design approaches. Thus, once explicit consent is gained, the consumer can be tracked on all websites, even if there are approaches to limit the privacy impact of such a decision (such as anonymization, truncation of URLs, limits on data retention, limits on secondary use) while still giving the consumer the benefit of the technology. This all or nothing approach fails to protect consumers regardless of the choices they take.

The self-regulatory programs are also deficient with regard to their accountability programs. Their compliance reviews are inadequate, they fail to set clear thresholds for sanctions, and they provide inadequate statistical reporting requirements to allow the public to monitor compliance. This represents a missed opportunity, because enforcement can be used as a tool to benefit advertisers and consumers by giving consumers privacy protections in exchange for business goodwill that will follow trustworthy practices.

Other aspects of the self-regulatory programs that are too numerous to mention in this proposal lend themselves well to critique. Our project will explore the legitimacy of these programs by scrutinizing their terms, evaluating the extent they meet consumer needs, and suggesting how to revise them to improve their efficacy.

#### **IV. Issues and Questions for Technologists**

*A. Does the name "Online Behavioral Advertising" adequately address policy concerns? Is it technically accurate to describe practices?*

The name of the practice under discussion has been debated at least several times in its history. In 2000, the Federal Trade Commission ("FTC") referred to the practice as "Online Profiling."<sup>7</sup> A 2009 FTC staff report now refers to the practice as "Online Behavioral

---

<sup>6</sup> National Consumer Council, *Models of self-regulation: An overview of models in business and the professions* 23 (November 2000), available at

[http://www.talkingcure.co.uk/articles/ncc\\_models\\_self\\_regulation.pdf](http://www.talkingcure.co.uk/articles/ncc_models_self_regulation.pdf) ("As far as practicable, the operation and control of the scheme should be separate from the institutions of the industry").

<sup>7</sup> "Online Profiling: A Report to Congress; Part 2: Recommendations," Federal Trade Commission, July 2000.

Advertising,”<sup>8</sup> which is also the industry’s preference.<sup>9</sup> The industry name for the practice is insufficient to embrace consumer concerns because it does not address the technological scope of the practice across multiple mediums<sup>10</sup> or the temporal scope that embodies the life of the data through collection, use, and storage.

Our project refers to the practice under discussion as “Consumer Tracking and Profiling” because consumers’ interests should be preserved even after advertisers gain explicit consent to use consumer data. “Tracking and Profiling” embodies the meaningful control consumers require over their data after it is collected.

The industry automatically comes out ahead with the word “advertising” embedded in the phrase OBA because consumers are familiar with the concept of advertising and they value the services supported by advertising. However, this practice implicates privacy concerns beyond the traditional sense of advertising. Consumers may not know that seemingly anonymous data could be linked to their personal identities.<sup>11</sup> In addition, consumers’ data could be sold to another entity for purposes outside of advertising if the advertising network folds or merges in the absence of clear self-regulatory principles against doing so. For example, with the prominent use of online job applications, employers could in theory buy this data from advertising networks to screen candidates based on behavioral information.

The phrase OBA also does not indicate the level of knowledge consumers might have about what information advertisers have on them. Wiretapping or an early 20th century telephone party line may be a more apt comparison to understand this aspect of the practice. An advertising network may have picked up a receiver, so to speak, unbeknownst to the consumer browsing the Internet, who merely thinks he is communicating with the website owner. This advertising network does not only listen in to the conversation, but he may conference in others parties and record the conversation for his benefit.

“Consumer Tracking and Profiling” is just one proposal we have in an attempt to more accurately describe such practice. We welcome suggestions from technologists at the April conference for proposals on the framing of this practice.

---

<sup>8</sup> “Self-Regulatory Principles For Online Behavioral Advertising,” Federal Trade Commission Staff Report, February 2009.

<sup>9</sup> NAI 2008, pg. 4, available at:

[http://www.networkadvertising.org/networks/2008%20NAI%20Principles\\_final%20for%20Website.pdf](http://www.networkadvertising.org/networks/2008%20NAI%20Principles_final%20for%20Website.pdf);

Self-Regulatory Principles for Online Behavioral Advertising, pg. 10, available at:

<http://www.aboutads.info/resource/download/seven-principles-07-01-09.pdf>.

<sup>10</sup> Including through applications that are not browsers, such as chat, and through other platforms, such as video game consoles.

<sup>11</sup> Testimony of Ashkan Soltani Before the Senate Committee on Commerce, Science, and Transportation Hearing on The State of Online Consumer Privacy, March 16, 2011 (“Despite some claims that these collected browsing profiles are “anonymous,” recent computer science research suggests that it is often quite easy to re-identify datasets that contain user information.”)

*B. Even when users opt out, advertisers may still track them and deliver ads. How will this affect the public perception of the Do Not Track Header?*

The self-regulatory principles under discussion limit a user's ability to opt-out. Consumers can only choose not to have their information exploited for online advertising. However, advertising networks could still collect data on all users, whether they opt-out of tracking or not. When a consumer opts-out, the data collected could be used for other purposes, so long as it is not behavioral advertising.

What do consumers expect to happen when they opt-out of the practice or alternatively, implement the Do Not Track Header on their browser? Do these self-regulatory principles undermine the expectation of no tracking from consumers? If the data collected is not used for behavioral advertising, what other purposes could they be used? Are those purposes necessary from a technical standpoint? Is there a way to technologically meet consumers' expectations on the collection of data?

*C. Data retention is permitted by NAI and DAA principles "as long as necessary to fulfill a legitimate business need." How long do network operators really need the data?*

The self-regulatory principles are extremely permissive in that they give businesses the ability to keep their data for essentially an unrestricted period of time. This is so because what constitutes a "legitimate business need" is left undefined. It is unclear how long businesses need to keep data to conduct behavioral advertising. How long is the data useful for the purposes of behavioral advertising? Are there diminishing returns, and are these quantifiable such they could inform a cost-benefit analysis?

*D. The NAI employs four staff members, one of which monitors the compliance of its 66 members.<sup>12</sup> From a technical perspective, what would be required to monitor compliance? Is one person sufficient to fulfill this task?*

Accountability measures should ensure that advertisers follow through on what they say they will do, e.g., stop behavioral advertising when a consumer opts out. Our white paper evaluates whether an appropriate amount of accountability exists in the self-regulation principles and will demonstrate that more can be done to ensure the integrity the programs. The question remains, however, what is entailed in evaluating compliance for dozens of advertising networks from a technical standpoint? Is one person enough to accomplish this task? What, in a technical sense, should accountability programs entail?

## **Conclusion**

We look forward to discussing these issues with workshop participants and believe that there is a great opportunity for knowledge mobilization among technologists and lawyers in the challenge to meaningfully evaluate self-regulatory programs for online advertising.

---

<sup>12</sup> <http://www.networkadvertising.org/about/staff.asp>

# Intel's Interest in W3C Tracking and Privacy Workshop

Narm Gadiraju: Systems Architect, Digital Home Group, Intel Corporation

## Participant's interest

Intel Corporation, a world leader in silicon innovation, develops technologies, products and initiatives to continually advance how people work and live. As a participating member of the consumer electronics (CE) ecosystem, Intel is interested in helping CE OEMs, content providers and service providers to bring the richness of the Internet to Television. In support of that goal, Intel is working with industry leaders to enable Smart TV experiences that go far beyond traditional Internet-connected consumer electronics devices. Smart TV helps consumers enjoy a virtually limitless array of Internet content, broadcast programming, personal media and a range of applications, all available on a single TV screen. From a silicon perspective, Intel has developed a line of system-on-a-chip (SoC) products targeted to digital TVs, optical media players and advanced set-top boxes, all of which are optimized for bringing internet content and applications to TV. Intel is interested in collaboratively working in the W3C to enable web standards that will accelerate the market adoption of a truly connected, immersive and 'smart' TV experience.

As a supplier of silicon products to both the IT and CE industry, Intel brings an exclusive viewpoint and technical competence in developing, enabling, and promoting robust platforms for the environments that W3C's future TV group is targeting. In W3C, Intel already participates in a number of HTML related Working Groups, such as the HTML WG, the Web Applications WG and the Device API and Policy WG. Furthermore, Intel joined the recent W3C's Web and TV interest group to address the requirements of the smart TV.

## Point of View

The HTML5 suite of specifications creates exciting new opportunities to bring the power and opportunities of the Web to new devices. Intel is appreciative of the W3C's efforts in organizing this workshop on Tracking and Privacy. As more internet content becomes available to users through connected TVs, personalizing browsing experience on TV to the user/family's viewing habits and interests is a key to the success of smart TV devices. However, when tracking is used as a tool to collect viewing information, it is critical that we ensure proper mechanisms are available to protect the privacy of the user. The privacy requirements for a 10 foot (TV) user experience are likely to be different from 2 foot (PC, SmartPhone, Tablet) user experience. We are interested to learn and contribute to the W3C's efforts to analyze this subject and create appropriate solutions.

## Position Paper for W3C Workshop on Web Tracking and User Privacy

William McGeeveran  
Associate Professor, University of Minnesota Law School

### Summary

Individual entities have begun to offer privacy-enhancing technological measures such as “do not track” browser extensions, and organizations such as W3C and the IETF are moving to consider systematic responses to privacy concerns. These are positive developments for improving user privacy, but I think they are unlikely to recognize their full potential if they are implemented in a legal vacuum. I want to explore the types of statutory and regulatory rules most likely to promote—but not interfere with—development of code-based privacy enhancements.

### Background

I am an associate professor at the University of Minnesota Law School. As a legal scholar I specialize in information law, particularly internet, privacy, and intellectual property issues. My privacy-related research focuses on the interaction between marketing practices, online technology, and legal rules.

In 2002, I wrote one of the earliest and most comprehensive examinations of P3P and its interaction with law, later published in the *NEW YORK UNIVERSITY LAW REVIEW*. (See William McGeeveran, *Programmed Privacy Promises: P3P and Web Privacy Law*, 76 *N.Y.U. L. REV.* 1812 (2002)) I concluded that light-touch legal regulation could have encouraged broader adoption of P3P and more broadly provided the necessary “nudge” to stimulate development of privacy-enhancing technological solutions. In a more recent article, published in the *UNIVERSITY OF ILLINOIS LAW REVIEW*, I analyzed the emerging practice of social marketing—defined there as the disclosure of an individual’s browsing and purchasing habits as a form of online word-of-mouth promotion aimed at that individual’s social network. (See William McGeeveran, *Disclosure, Endorsement, and Identity in Social Marketing*, 2009 *U. ILL. L. REV.* 1105) This piece reached a conclusion quite similar to my P3P analysis: a more robust legal requirement for user consent would stimulate market and technological best practices to shape the emerging field of social marketing in a manner that protects privacy. As I argued in both articles, that respect for privacy not only helps individual users, but it safeguards the economic and communicative vitality of the internet. Online advertisement or the recommendation ecology are

compromised by user suspicion and the potential “spammification” of inaccurate, exaggerated, or undesired disclosures about individual preferences.

I have begun research exploring application of a similar legal analysis to the general topics of behaviorally-based marketing. (In a related vein, I am also considering privacy implications of various identity-layer proposals.)

## **Position**

Twelve years after the publication of Lawrence Lessig’s *Code and Other Laws of Cyberspace*, his descriptive observations about the application of architecture, law, markets, and norms to online behavior have become so widely internalized that they border on cliché. Yet the prescriptive analysis that he began there, and that has been continued in work by many others academics, technologists, and activists, has not achieved the same level of acceptance.

Unfortunately, some policymakers still tend to propose rigid mandates, and in response some in the internet community tend to view all regulatory interventions with hostility. Lessig would argue that the modality of traditional law should be used to shape the development of other modalities to achieve goals such as enhanced user privacy. This has become a common mode of regulation in, for example, environmental law.

Current legal rules concerning commercial data-handling (or the lack of them) have failed. Without channeling from legal rules, the general and diffuse consumer demand for increased online privacy has not coalesced into a coherent demand for consistent treatment (beyond occasional media-fueled uproars over particular practices) or created momentum for particular privacy-protecting technology. But this does not mean that users are satisfied with the level of privacy they now enjoy, particularly when they learn more about current practices. We are all familiar with empirical evidence telling us so.

A legal regime that imposed certain minimum requirements and mandated respect for certain expressed user preferences could foster widespread adoption of uniform best practices—and the technological tools for achieving them. W3C obviously would play a crucial role in fostering such an environment. I continue to believe that, had such legal rules been in place ten years ago, the development and adoption of P3P might have turned out differently.

I hope my participation in the workshop will help inform analysis of legal rules calculated to support and promote, rather than impede, robust technological response to privacy concerns.

---

# Toward Privacy Standards Based on Empirical Studies

**Serge Egelman and Erika McCallister**

National Institute of Standards and Technology  
100 Bureau Drive  
Gaithersburg, MD 20899  
{serge.egelman, erika.mccallister}@nist.gov

## Abstract

In this paper, we argue that if privacy standards are created to guide “do-not-track” technologies, these standards should be created with the primary stakeholder in mind: the data subject. Previous privacy and security standards have been unsuccessful because implementations were inconsistent, confusing, or not readily apparent to the user. The Fair Information Practice Principles (FIPPs) empower users to make informed decisions about their privacy and should be the basis for any resulting privacy standard. However, research must be conducted to determine best practices for presenting this information to users. We describe one such study that we are currently conducting and what we expect to learn about promoting informed consent with regard to data sharing.

## Keywords

Privacy standards, empirical studies, informed consent

## Introduction

The US Department of Commerce recently released its Privacy Green Paper [17], which made recommendations for the future of Internet privacy.

The Department sought to balance consumer trust with commercial innovation. Among its recommendations was the idea of using enforceable codes of conduct based on a set of Fair Information Practice Principles (FIPPs). The principles of transparency and individual participation are directly relevant to the concept of do-not-track. Transparency means that organizations should notify individuals regarding the collection, use, dissemination, and maintenance of their personal information. Individual participation means that organizations should provide means of consent to individuals regarding the collection, use, dissemination, and maintenance of their personal information, as well as provide a means for access to and correction of personal information. Users must know what information is collected and how it is used to make good decisions about when to use a do-not-track technology.

## Background

Studies of user perceptions of privacy have found that Internet users are generally concerned about their privacy when online [1]. However, studies have found that in practice their actions do not reflect their preferences [2, 3, 4]. Part of the problem is that users are often unaware of the types and amounts of information that are shared with affiliated websites, which websites are affiliated, or how they can opt-out of having their information shared [13]. However,

when given effective privacy tools with which they can state their privacy preferences, observed behaviors become better aligned with stated preferences [8, 16]. Such privacy tools are effective because the user is at the center of the design and empirical data on user behavior informs the design decisions [20].

Many current Internet privacy tools do not adequately allow users to make informed decisions because the standards on which they are based have not incorporated empirical data on how to best support users' needs. For instance, studies have shown that many web browser security indicators go unnoticed because the indicators are outside the user's view [18, 19], are inconsistent across vendors and versions [14], and have unclear meanings [9].

The W3C's P3P standard attempted to empower users to make informed privacy decisions by specifying a format in which websites could post machine-readable privacy policies [5], but left it up to software vendors to determine what information to display to users and how it was to be displayed. Despite research showing P3P adoption rates of over 25% on popular websites [7, 10], use of full P3P policies failed to gain traction.<sup>1</sup> This may be due in part to browser-based P3P implementations that were hard to understand and often went unnoticed [6]. The onus of this failure is not necessarily the fault of software developers or designers. P3P is a comprehensive standard in its focus on converting natural language privacy policies into a machine-readable format. However, P3P lacks what

---

<sup>1</sup> The P3P compact policy is widely used today, but use of the full XML policy never reached the level that its creators expected [15].

has proven to be essential guidance on parsing this detailed information in a way that will allow users to take action. Research has now shown that P3P implementations could be designed to help users make more informed choices (e.g., [8, 11]), but it is likely too late to update the standard at this point, and seems unlikely to gain sufficient traction in the future.

In order to be successful, technical standards used to assist in the implementation of the FIPPs must be objective and based on empirical evidence. Since the FIPPs rely on users being able to provide informed consent for data sharing activities, technical standards need to specify how to effectively communicate privacy information to users. At NIST, we are in the early stages of conducting a study to determine effective interfaces for obtaining informed consent for websites to share data with affiliates. In the next section, we provide an overview of this study in order to show how empirical studies can better inform privacy standards.

### **Study Design and Goals**

We have designed a study to examine how participants' data-sharing decisions change based on the presence of salient information describing the data to be shared. Specifically, we are examining a popular single sign-on (SSO) interface to examine whether participants make different decisions about whether to use SSO based on how the data being shared is described. Many websites are opting to support various SSO platforms because in addition to simplifying their authentication implementations, it allows these websites to collect more data about their users and their users' habits. Providing informed consent in this situation goes to the very heart of "transparency." While our study specifically examines privacy trade-offs when using an



SSO implementation, the results should be generalizable to the design of any dialog used to solicit informed consent to collect or share personal information from users. Specifically, we expect these results to be relevant to the design of a granular do-not-track interface that allows users to opt-in or opt-out of tracking based on the requesting entity and the data requested.

The particular SSO implementation that we are studying was developed by a social networking website. In addition to offering affiliated websites the ability to authenticate users, the social networking site also provides the affiliated websites with personal information from users' profiles. This data may be used for marketing, user profiling, or other unknown reasons. However, the user must first consent to sharing this data. Figure 1 depicts a screenshot of the original consent dialog. As can be seen, the dialog ostensibly supports some of the FIPPs by providing a list of data being requested and the name of the organization requesting it. In this particular example, the dialog is requesting:

- Name
- Profile picture
- Gender
- Networks
- User ID
- List of friends
- Any other public profile information

We are examining whether the information is presented clearly enough to facilitate informed consent. Specifically, we have created an experimental condition (Figure 2) wherein users instead see their data verbatim, in addition to the descriptions of that data. If we find that users in the experimental condition were

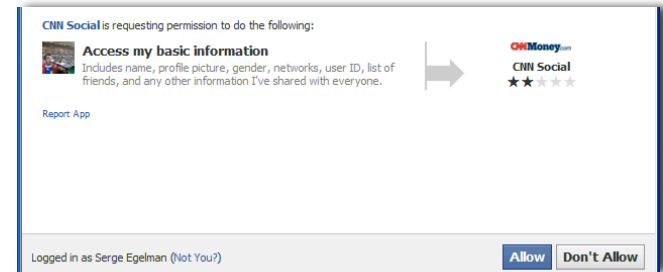


Figure 1: Screenshot of the consent dialog for sharing profile data with participating websites.



Figure 2: Screenshot of consent dialog in our experimental condition. Here, data to be shared is displayed verbatim.

significantly more or less likely to use the SSO option to authenticate to the same websites as users in the control condition, then it is likely because they better understood what data the websites were requesting. This will yield important guidance on how to better request informed consent from users such that it is truly informed.

## Conclusion

Our study aims to answer several questions about how to more effectively support the FIPPs through better user interaction design. At the workshop, we hope to present a larger set of questions related to the concept of do-not-track and how empirical studies will better

inform a technical standard. Some related standards proposals focus on the binary decision of track or no-track [12]. We propose the use of empirical research to create objective and usable standards that balance user privacy preferences, tailoring to users' needs, and the commercial innovation that can be gained through sharing user data. Do-not-track is not a binary question; it should be more granular by focusing on the ways of conveying information to the user. Do-not-track cannot be effective without the transparency created through granularity. Empirical research, such as our current study, should be used as input for any potential do-not-track standards to improve the usability, effectiveness, and adoption.

## References

- [1] Ackerman, M. S., Cranor, L. F., and Reagle, J. 1999. Privacy in e-commerce: examining user scenarios and privacy preferences. In *Proceedings of the 1st ACM conference on Electronic commerce* (EC '99). ACM, New York, NY, USA, 1-8.
- [2] Acquisti, A. 2004. Privacy in electronic commerce and the economics of immediate gratification. In *Proceedings of the 5th ACM conference on Electronic commerce* (EC '04). ACM, New York, NY, USA, 21-29.
- [3] Acquisti, A. and Grossklags, J. 2005. Privacy and Rationality in Individual Decision Making. *IEEE Security and Privacy* 3, 1 (January 2005), 26-33.
- [4] Berendt, B., Gunther, O., and Spiekermann, S. 2005. Privacy in e-commerce: stated preferences vs. actual behavior. *Commun. ACM* 48, 4 (April 2005), 101-106.
- [5] Cranor, L. 2002. Web Privacy with P3P. O'Reilly & Associates, Sebastopol, CA.
- [6] Cranor, L. F., Guduru, P., and Arjula, M. 2006. User interfaces for privacy agents. *ACM Trans. Comput.-Hum. Interact.* 13, 2 (June 2006), 135-178.
- [7] Egelman, S., Cranor, L. F., and Chowdhury, A. 2006. An analysis of P3P-enabled web sites among top-20 search results. In *Proceedings of the 8th International Conference on Electronic Commerce*. (ICEC '06). ACM, New York, NY, USA, 197-207.
- [8] Egelman, S., Tsai, J., Cranor, L. F., and Acquisti, A. 2009. Timing is everything?: the effects of timing and placement of online privacy indicators. In *Proceedings of the 27th international conference on Human factors in computing systems* (CHI '09). ACM, New York, NY, USA, 319-328.
- [9] Friedman, B., Hurley, D., Howe, D. C., Felten, E., and Nissenbaum, H. 2002. Users' conceptions of web security: a comparative study. In *CHI '02 extended abstracts on Human factors in computing systems* (CHI EA '02). ACM, New York, NY, USA, 746-747.
- [10] Jensen, C., Sarkar, C., Jensen, C., and Potts, C. 2007. Tracking website data-collection and privacy practices with the iWatch web crawler. In *Proceedings of the 3rd Symposium on Usable Privacy and Security* (SOUPS '07). ACM, New York, NY, USA, 29-40.
- [11] Kelley, P. G., Bresee, J., Cranor, L. F., and Reeder, R. W. 2009. A "nutrition label" for privacy. In *Proceedings of the 5th Symposium on Usable Privacy and Security* (SOUPS '09). ACM, New York, NY, USA, , Article 4 , 12 pages.
- [12] Mayer, J., Narayanan, A., and Stamm, S. "Do Not Track: A Universal Third-Party Web Tracking Opt Out." IETF draft document. Accessed: March 18, 2011. <http://tools.ietf.org/html/draft-mayer-do-not-track-00>.
- [13] McDonald, A. M. 2010. Cookie confusion: do browser interfaces undermine understanding?. In *Proceedings of the 28th of the international conference extended abstracts on Human factors in computing systems* (CHI EA '10). ACM, New York, NY, USA, 4393-4398.
- [14] Schultze, S. "Web Browser Security User Interfaces: Hard to Get Right and Increasingly

- Inconsistent." Accessed: March 17, 2011.  
[http://www.freedom-to-tinker.com/blog/sjs/web-browser-security-user-interfaces-hard-get-right-and-increasingly-inconsistent\\_](http://www.freedom-to-tinker.com/blog/sjs/web-browser-security-user-interfaces-hard-get-right-and-increasingly-inconsistent_)
- [15] Schwartz, A. "Looking Back at P3P: Lessons for the Future." November 11, 2009.  
<http://www.cdt.org/paper/looking-back-p3p-lessons-future>.
- [16] Tsai, J., Egelman, S., Cranor, L., and Acquisti, A. The effect of online privacy information on purchasing behavior: An experimental study. In *Proceedings of the 2007 Workshop on the Economics of Information Security (WEIS'07)* (Pittsburgh, PA, USA, 2007).
- [17] US Department of Commerce. "Commercial Data Privacy and Innovation in The Internet Economy: A Dynamic Policy Framework." December 2010.  
<http://www.commerce.gov/sites/default/files/documents/2010/december/iptf-privacy-green-paper.pdf>.
- [18] Whalen, T. and Inkpen, K. M. 2005. Gathering evidence: use of visual security cues in web browsers. In *Proceedings of Graphics Interface 2005* (GI '05). Canadian Human-Computer Communications Society, School of Computer Science, University of Waterloo, Waterloo, Ontario, Canada, 137-144.
- [19] Wu, M., Miller, R. C., and Garfinkel, S. L. 2006. Do security toolbars actually prevent phishing attacks?. In *Proceedings of the SIGCHI conference on Human Factors in computing systems* (CHI '06), Rebecca Grinter, Thomas Rodden, Paul Aoki, Ed Cutrell, Robin Jeffries, and Gary Olson (Eds.). ACM, New York, NY, USA, 601-610.
- [20] Zurko, M. E. and Simon, R. T. 1996. User-centered security. In *Proceedings of the 1996 New Security Paradigms Workshop* (NSPW '96). ACM, New York, NY, USA, 27-33.

# Content Based Do Not Track mechanism

**Vincent Toubiana**

v.toubiana@free.fr

**Helen Nissenbaum**

Helen.Nissenbaum@nyu.edu

## 1 Problem

There is currently a debate about the form that a DNT implementation should have to provide users with a clear control over the data they ‘provide’ to tracking ad-network. Currently, all implementations of the DNT concept do not consider the content of the visited web page. Once the user opted-out from tracking, his decision have an impact on every website he surfs on unless he opt-back. The idea emerged to use a DOM flag to inform websites of the user preference [1].

Current approaches suggest that user decision to opt-back should be expressed at a domain or website level. While these approaches provide user ability to opt-back for selected websites, they present the following drawbacks:

- users can only opt-back to a domain or website after they visit it, so unknown websites are ‘not-authorized’ by default,
- opt-back decision remains even if the website content change (this is particularly problematic for news websites) even when the domain name is transferred to another entity.

Consequently, if user a visits websites very occasionally (for instance once in a month), he is very unlikely to opt-back for that website even if he does not mind being tracked during these visits.

In addition, each approach presents some drawbacks that could limit its adoption. On the one hand, a list of ‘authorized’ websites could contain too many entries and would become. On the other hand, a domain list would be easier to manage but may not accurately reflect user’s decisions<sup>1</sup>.

## 2 Solution

With regard to the drawbacks that existing approaches present, we propose an alternative approach where users opt-back to ‘topic’ rather than website. Our approach provides users with a simple control over the information that can be inferred from their browsing habits without compromising their privacy. With this solution users could specify on which category of website they agree to be tracked. A website category would be determined from content analysis and if the users accepted to be tracked on this category, then no DNT mechanism would be used when downloading third-party ads and trackers published on this website.

When the website belongs to a ‘not trackable’ category, then the usual tracking prevention mechanism will be employed when downloading ads and trackers. In fact users could even manage several ‘tracked’ profiles, each of them containing different categories (with no overlap) and being associated to different cookies (see Figure 1). In that situation, there should be no possibility for ad-networks to link the two profiles that would be seen though two different cookies (linking based on the IP address should be prevented by policy).

---

<sup>1</sup> J. Mayer on twitter: “Another example of why domain names aren't the right privacy boundaries: [metrics.apple.com](https://metrics.apple.com) = Adobe (formerly Omniture)”

Categories	Profile P1	Profile P2	Do Not Track
Arts & Entertainment	●		
Autos & Vehicles	●		
Computers & Electronics		●	
Finance			●
Internet & Telecom		●	
Law & Government			●
People & Society	●		
Sports		●	
Travel	●		
Cookies ID			
DoubleClick	123-DC-XY	456-DC-AB	DNT Header
Microsoft	789-MS	ABC-MS	DNT Header

Figure 1: Profile configuration and mapping with cookies

In this approach, we could reuse the set of categories adopted by ad-networks [2][3] (and based on ODP [4]) to categorize websites and facilitate the migration of users who have already been tracked and profiled.

Notice that focusing on the top-level categories defined by those ad-networks could be enough to provide users with a good control over where they can be tracked.

### 3 Implementation

This section provides guidelines to realize it either has a pure browser extension or as a system supported by both browsers and ad-networks.

When loading a new website, the browser determines the website topic either by calling an internal routine or by using information provided by a third party (eventually an ad-network). Once the website main topic is identified, the browser retrieves the profile (shaped as a list of categories) related to that topic. If a profile is found, the browser loads it and set the corresponding cookies when sending requests to ad-networks and other trackers. If no profile contains that topic, the browser enters in Do Not Track mode when sending request to the ad-networks publishing ads and trackers on the website.

Here we assume that ad-networks adopt the list of categories of Google Ad Preference manager [2]. This list of categories, like the one proposed by Yahoo! [3], is based on ODP categorization [4]. While an agreement on the categorization used by the different ad-networks is not essential for our solution to work, it would help evaluating the accuracy of the employed categorization algorithm if users could verify on each ad-network page that the profile linked to each of their cookies match their expectation.

#### 3.1 Browser implementation

We would implement our solution in a browser using an embedded categorizer like the one used in Adnostic [4]. This categorizer uses information provided in the metadata of webpage's to determine the page's content and identify its main topic. The extension would then set the appropriate headers and cookies before starting to download ads and trackers. The topic corresponding to the page would be kept in cache and would be re-evaluated after cache expiration or when the metadata changes. Figure 2 illustrates this process.

Furthermore, users could edit and share the established list of correspondences between websites and topics.

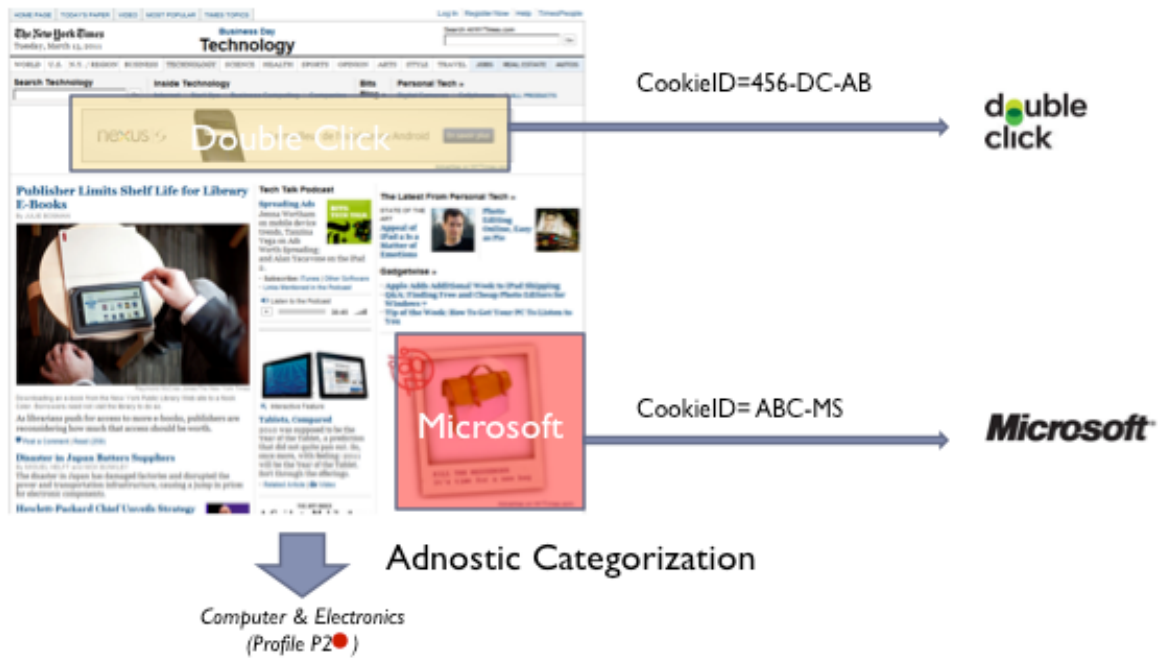


Figure 2 Pure browser solution description. The category is first determined (Computer and Electronics) and mapped to a profile. Then ads are downloaded with the right profile.

### 3.2 Ad-network supported implementation

An alternative would be to offload the categorization process to ad-networks who would inform users of the category corresponding to a visited website. This category is identified by ad-networks to place contextual ads and could be either published and sent to the user ‘offline’ (like Google safe-browsing) or sent when she makes her first request to the ad-network from an uncategorized website (there are often several requests sent by the browser to download ads). The first request could be sent with a DNT header and, when sending the next request, the header would be set according to the corresponding user’s profile.

The categories list provided by ad-network could be verified by checking that contextual ads displayed on a website are related to the list of categories provided. Another incentive for ad-networks to provide web site categorization is that it’ll reduce the rendering time of their ads for users who opted-back to some categories.

For our approach to be the most effective, a similar categorization could be used by every ad-network, thus limiting the number of required requests to one per page, independently of the number of ad-networks publishing on it.

## References

- [1] A. Cooper and H. Tschofenig, “Overview of Universal Opt-Out Mechanisms for Web Tracking”
- [2] Google Ads Preference Manager, <http://www.google.com/ads/preferences>
- [3] Yahoo! Ads Interest Manager, [http://info.yahoo.com/privacy/us/yahoo/opt\\_out/targeting/](http://info.yahoo.com/privacy/us/yahoo/opt_out/targeting/)
- [4] DMOZ, Open Directory Project, <http://www.dmoz.org/>
- [5] V.Toubiana, A. Narayanan, D. Boneh, H. Nissenbaum, and S.Barocas, “Adnostic: Privacy Preserving Targeted Advertising”, In Proceedings of NDSS 2010

# Do Not Track

*Nokia Browser Position*

- Vikram Malaiya

# Our understanding of Do Not Track (DNT)

---

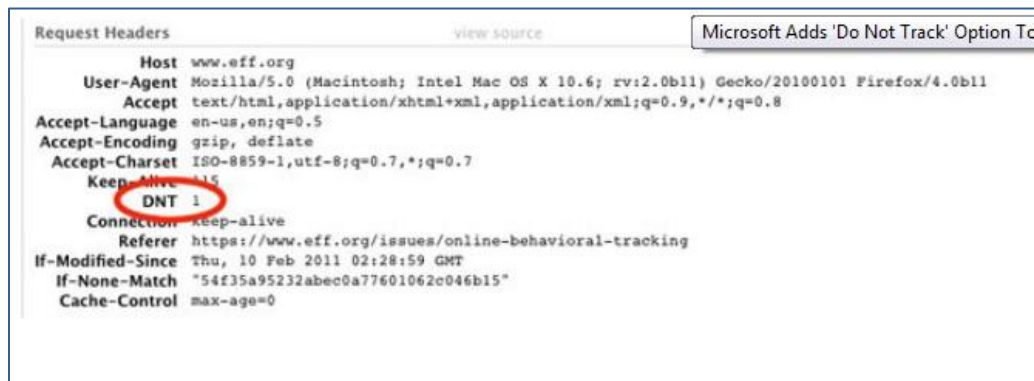
- DNT is a technology to enables users to opt out of third-party web tracking
- No agreed upon definition of DNT. There are currently 3 major technology proposals for responding to third-party privacy concern.
  1. Stanford University and Mozilla's DNT HTTP Header technique.
  2. Blacklist based technique such as Microsoft's 'Tracking Protection' which is part of IE9
  3. Network Advertising Initiative's model of a per company opt-out cookie. Opt-out cookie approach is being promoted by Google.



# DNT as HTTP Header

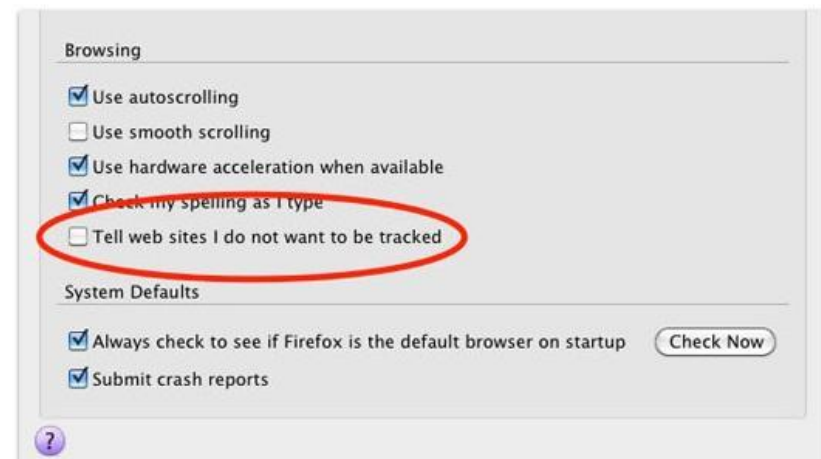
The Browser adds 'DNT' / 'X-Do-Not-Track' to its http header. The header is sent out to the server with every web request. This header acts as a signal to the server suggesting that the user wishes to opt out of tracking.

Adoption: Firefox 4, IE9



Request Headers [view source](#) Microsoft Adds 'Do Not Track' Option To

```
Host: www.eff.org
User-Agent: Mozilla/5.0 (Macintosh; Intel Mac OS X 10.6; rv:2.0b11) Gecko/20100101 Firefox/4.0b11
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip, deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 115
DNT: 1
Connection: keep-alive
Referer: https://www.eff.org/issues/online-behavioral-tracking
If-Modified-Since: Thu, 10 Feb 2011 02:28:59 GMT
If-None-Match: "54f35a95232abec0a77601062c046b15"
Cache-Control: max-age=0
```



Browsing

- Use autoscrolling
- Use smooth scrolling
- Use hardware acceleration when available
- Check my spelling as I type
- Tell web sites I do not want to be tracked

System Defaults

- Always check to see if Firefox is the default browser on startup [Check Now](#)
- Submit crash reports

# DNT as HTTP Header

---

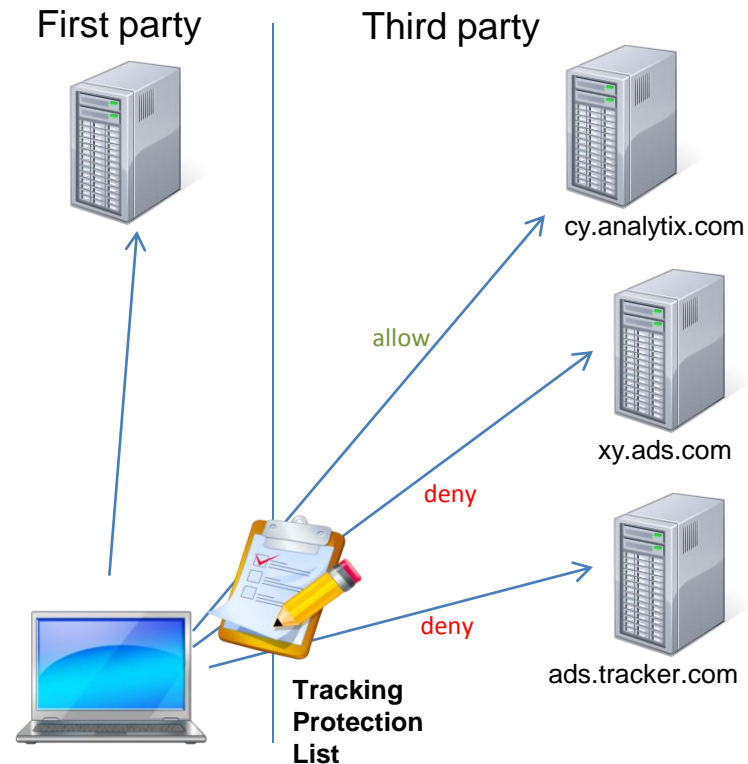
- Pros:
  - Scope: Server could apply restrictions to all third party entities and tracking mechanisms
  - Persistent: No reconfiguration needed once set
  - Simple: Easy to implement on the browser side
- Cons:
  - Only work as long as the server honors users preferences
  - No way to enforce national regulations/legislations to servers located beyond country boundaries

# Block(Black) List / Tracking Protection

This is a consumer opt-in mechanism which blocks web connections from known tracking domains that are compiled on a list.

Adoption: 'Tracking Protection'  
in Internet Explorer 9

The downloadable Tracking Protection Lists enable IE9 consumers to control what third-party site content can track them when they're online.



# Block(Black) List / Tracking Protection

---

- Pros:
  - More reliable than http header, because it put no reliance on trusting the server to honor user preferences, and it transcends national legal boundaries
  - Blocks third-party cookies, tracking pixels, web beacons, hit counters, analytics scripts, and other tools used for tracking.
  - Blocks ads as well (Pro/Con)
- Cons:
  - Only covers resources on the block list
  - Consumers have to judge the merit of a block list
  - Block lists need to be actively updated
  - Big players such as Google/Facebook not on the block list would still be able to track user behavior as a third party

# Opt-Outs Cookie Approach

---

- An Icon based self-regulatory approach proposed by Network Advertising Initiative (NAI) in US, and by European advertising industry alliance(EASA) in EU
- The scheme does not depend on any special Browser setting, it works by adding an icon to behavioral ads served on websites to indicate it is a behavioral ad
- A click on the icon leads the user to [www.youonlinechoices.com](http://www.youonlinechoices.com) (EU), or [www.aboutads.info](http://www.aboutads.info) (US). These websites allow users to opt out of behavioral advertising by selecting one or all advertisers that are listed as serving him behavioral ads
- The sites set a third party (opt-out) cookie on user's browser to capture's his choice. This cookie goes out to the advertisers in the subsequent browser sessions to indicate user's choice

Adoption: Google Chrome's Keep My Opt-Outs Extension, helps user maintain persistent opt-out cookie



# Cookie Opt-Outs Approach

---

- Pros:
  - Driven by a industry driven self regulatory program
- Cons:
  - Lack of icon visibility, poor icon placement will render this approach ineffective
  - Persistence, not clear is the cookies could be accidentally deleted
  - Narrowly focused on only online advertisements
  - Only covers the ~70 NAI members in US
  - No visibility into commitment of participating advertisers. Advertiser could choose to honor user's request based on their commitment/compliance to NAI/EASA's best practices recommendations

# Position

---

- The HTTP header based DNT approach has merits because the simplicity and built in persistence in its design. However, given the cons mentioned in this report, this scheme alone may not be enough to protect online privacy, but it is a good step forward
- This scheme would complement the EU privacy directive that calls for “explicit consent” to be collected from Internet users who are being tracked via cookies. This directive comes in effect in May 2011
- We also support the self-regulatory opt-outs approach proposed by NAI and EASA, however this approach needs to resolve the open questions that we have posed in this paper to be effective. Moreover, such approach requires wider adoption by companies across Globe

## **Submission to the W3C Workshop on Web Tracking and User Privacy 28/29 April 2011, Princeton, NJ, USA**

By: The Personal Data Ecosystem Consortium

Kaliya Hamlin, Executive Director, Personal Data Ecosystem Collaborative Consortium  
[director@personaldataecosystem.org](mailto:director@personaldataecosystem.org) @identitywoman Mobile: 510-472-9069

Mary Hodder, Chairman, Personal Data Ecosystem Collaborative Consortium  
[mary@personaldataecosystem.org](mailto:mary@personaldataecosystem.org) Mobile: 510-701-1975

Submitted April 1, 2011

*I just learned about this workshop at a dinner yesterday and I am submitting the position paper accordingly. I believe view we represent is vital to bring into the discussion as a better approach to user privacy. I hope you will still consider our position for inclusion in the workshop.*

### **Personal Data Storage and Services: A Middle Way between Do Not Track and Business as Usual Tracking**

The Personal Data Ecosystem Consortium represents a community of end-user advocates and technology innovators focused on individual rights and access to individuals' own personal data, and the business and innovation opportunity that this new user-management and control.

The perspective that we have is quite different than either Do Not Track or Business as Usual Tracking. Personal Data Storage and Services where individuals effectively "stalk themselves" aggregating their own data streams from diverse sources into a personal data service.

The scope of activity in this space goes beyond web surfing habits and tracking data. We chose to reflect this broader scope in this position paper because we know services are being built beyond that scope there are over 20 startups are building today for this emerging ecosystem and large firms are doing incubator projects testing the waters and the World Economic Forum just released a report articulating the value of Personal Data as an Emerging Asset Class.

We believe this approach has great promise because besides empowering users it also represents new business opportunities and incentives for companies who currently are dependent on the information they glean from a whole range of online tracking techniques to have access to even better information about the people they are seeking to market to.

#### **The Two Ends of the Spectrum**

On one end of the spectrum is the "Do not track" view, which relies on using technology and a legal mandate to prevent any data collection (as per the FTC Proposal). In this scenario, cross site behavioral targeting is suppressed because users signal they do not want any information to be collected on them as they move about the web. In this approach the economic value advertisers have been getting through higher click-through rates by providing more targeted ads is eliminated and sites that receive revenue from serving targeted ads is reduced if not eliminated. The economic value of the data is not captured by the end-user nor is it benefiting the media/advertising/data aggregating complex.

On the other end of the spectrum is the mode where we leave "Business as usual" in place as it has developed over the last few years. The door is wide open for ever more "innovative" pervasive and intrusive data collection, tracking and cross referencing for behavioral targeting in developing profiles -- digital dossiers created on billions of people, without their knowledge or consent, based on IP address, device identification, e-mail address etc. The status quo is highly invasive of people's privacy, linking their activities across contexts they wish to keep separate or private if they chose to do so. In addition, decisions about people's lives are increasingly made from such data, and they are not aware of it, though



the consequences can be quite severe. Economic value is derived, but at the expense of the basic dignity and privacy rights (ie personal control) of the individual.

## **Personal Data Storage and Services**

Personal data storage services are emerging, representing a middle way through, to provide an opt-in modality with greater choice and control to the individual over their data AND offer greater economic value to the business community, with huge innovation and market opportunities. This market, we believe, will be much larger than the current one based upon surreptitious stalking, and be based upon an ethical model involving the user in the transactions the might occur with their data, where choice, transparency, access and control are central features for users.

As envisioned, Personal Data Storage Services (PDS) allow individuals to aggregate their personal data, to manage it and then give permissioned access to businesses and services they choose -- businesses they trust to provide better customization, transparency, access and the ability to correct, as well more relevant search results and commercial offers, resulting in increased value for the user from their data.

Over the last year, activity in this space has grown tremendously. In this emerging field of innovation, we have identified over thirteen startups (some of them with significant venture capital funding), at least three open source projects, several technical standards efforts in recognized international standards organizations along with companies in the web, mobile, entertainment and banking industries working on this model.

One of the most important things about this emerging space is that it has engendered active business development both in the United States and across Europe. In other words, this model is viable across North American and European privacy regimes. Furthermore, the PDS model offers the possibility of achieving global interoperability, one of the key goals articulated by the Commerce Department for this forthcoming set of policies and regulations.

## **People are the Only Ethical Integration Point for Disparate Data Sets**

Today there is a personal data ecosystem emerging in which almost everyone unknowingly participates but without the personal individual controls to afford user-centric privacy. People unwittingly emit information about themselves, their activities and intentions, in various digital forms. It is collected by a wide range of institutions and businesses with which people interact directly; then it is assembled by data brokers and sold to data users (ie businesses that exploit our data without including us in the transaction). This chain of activity happens with almost no participation or awareness on the part of the data subject: the individual.

We believe that the individual is the only ethical integration point for this comprehensive and vast range of disparate personal data. For example, the list of data types below was put together by Marc Davis for the World Economic Forum talk: Re-Thinking Personal Data event in June of 2010. It highlights the vast range of datasets about an individual that might be in some digital form in some database somewhere.

### **Identity and Relationships:**

- \* Identity (IDs, User Names, Email Addresses, Phone Numbers, Nicknames, Passwords, Personas)
- \* Demographic Data (Age, Sex, Addresses, Education, Work History, Resume)
- \* Interests (Declared Interests, Likes, Favorites, Tags, Preferences, Settings)
- \* Personal Devices (Device IDs, IP Addresses, Bluetooth IDs, SSIDs, SIMs, IMEIs, etc.)
- \* Relationships (Address Book Contacts, Communications Contacts, Social Network Relationships, Family Relationships and Genealogy, Group Memberships, Call Logs, Messaging Logs)

### **Context:**

- \* Location (Current Location, Past Locations, Planned Future Locations)
- \* People (Co-present and Interacted-with People in the World and on the Web)
- \* Objects (Co-present and Interacted-with Real World Objects)
- \* Events (Calendar Data, Event Data from Web Services)

### **Activity:**

- \* Browser Activity (Clicks, Keystrokes, Sites Visited, Queries, Bookmarks)
- \* Client Applications and OS Activity (Clicks, Keystrokes, Applications, OS Functions)

\* Real World Activity (Eating, Drinking, Driving, Shopping, Sleeping, etc.)

#### **Communications:**

- \* Text (SMS, IM, Email, Attachments, Direct Messages, Status Text, Shared Bookmarks, Shared Links Comments, Blog Posts, Documents)
- \* Speech (Voice Calls, Voice Mail)
- \* Social Media (Photos, Videos, Streamed Video, Podcasts, Produced Music, Software)
- \* Presence (Communication Availability and Channels)

#### **Content:**

- \* Private Documents (Word Processing Documents, Spreadsheets, Project Plans, Presentations, etc.)
- \* Consumed Media (Books, Photos, Videos, Music, Podcasts, Audiobooks, Games, Software)
- \* Financial Data (Income, Expenses, Transactions, Accounts, Assets, Liabilities, Insurance, Corporations, Taxes, Credit Rating)
- \* Digital Records of Physical Goods (Real Estate, Vehicles, Personal Effects)
- \* Virtual Goods (Objects, Gifts, Currencies)

#### **Health Data:**

- \* Health Care Data (Prescriptions, Medical Records, Genetic Code, Medical Device Data Logs)
- \* Health Insurance Data (Claims, Payments, Coverage)

#### **Other Institutional Data:**

- \* Governmental Data (Legal Names, Records of Birth, Marriage, Divorce, Death, Law Enforcement Records, Military Service)
- \* Academic Data (Exams, Student Projects, Transcripts, Degrees)
- \* Employer Data (Reviews, Actions, Promotions)

In addition to this list, there is also the emerging wellness, or "quantified self," data that some users are beginning to collect about themselves through life-tracking companies, including daily or more granular statistics about their bodies and wellness activities. There is travel data including miles, trips and future plans.

#### **'Service Providers Must Work For the End-User**

Most people do not host their own e-mail servers or websites on servers in their basements. Similarly, most individuals will not have the technical skill or desire to actually manage the collection, integration, analysis, permission management and other services needed to derive value from their data. However, the fact that a few users can host their own email means the open standards for email and http are available top to bottom. We want to see Personal Data Services available through open standards, open source code, and an ecosystem that will interact with people who host their own PDS.

But mostly, individuals need to be able to trust that service providers in the Personal Data Ecosystem are working on the user's behalf. Given the sensitivity of the data, and the complexity of running their own servers, most users will rely on Personal Data Service providers. In addition, market models need to emerge that support the Personal Data Store Service Provider making money while working on the users' behalf. The Personal Data Ecosystem Consortium has a Value Network Mapping and Analysis project to outline this model and is raising money to support and foster the model.

#### **Personal Data should be treated like Personal Money**

Individuals must be able to move data between service providers, as they can move money between banks, retaining its value. However, with user's data, it's the user that is the provider, but there must still be many takers because of open data formats, activity streaming, and clear identity models that are also portable and separate from the data bank.

End-user choice and the right to transfer data from one service provider to another is key to this model. Just as our money does not become worthless when we move it from one bank to another, the same needs to hold true for individuals' data.

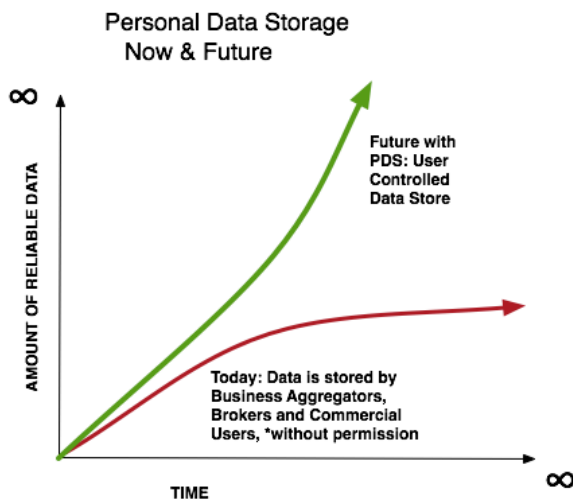
#### **Consumers need to be able to Collect and Aggregate Their Data from Product and Service Providers**

For this Personal Data Ecosystem and Economy to emerge and for user's to be properly protected, it is essential that users have easy access to their data from the providers with whom they do business. The



## Keeping our Data for a Lifetime, If We Want to do so

What if the individual could choose to retain all or a subset of the information about themselves for as long as they wanted? This is a graph that shows today's current data environment and a future where people are in control of their own data, and the opportunities around opt-in, more reliable data than stalking users surreptitiously currently permits.



The red line shows us what's happening today: some data aggregators are necessarily self-regulating by limiting the amount of time they keep data, and governments are limiting data retention and anonymization practices. And much data that is collected is without explicit permission, other than through onerous privacy policy the user agrees to once (usually) and the green line shows us what WOULD happen if people were given the capacity to store and manage their own data – if they could keep as much data as they wanted for as long as they wanted, or not at all, in their own data banks. Digital footprints reflecting a lifetime could be shared with future generations, people could self-assess, and applications through a marketplace would emerge to create new businesses and data uses we haven't yet thought of. In this user-centric model, the individual can aggregate information about themselves, where new classes of services more specific to the individual, based on data accessed with user permission, can emerge.

The foundation of this ecosystem is personal data storage services that are totally under the control of the individual. But a user-centric identity system needs to function in partnership with it (separate from a PDS) and we will need a regulatory regime that supports both of these technology solutions in user-centric form, where users own and control their own data.

This model where individuals are in control of their own data, aligns with the interests of all the stakeholders that we are seeking to balance. Only the data brokers and aggregators lose and they could refactor to have new roles in this ecosystem of end-user controlled data.

**Companies who collect personal data win.** By sharing and synchronizing with people's personal data stores, companies get far more accurate information. New services can be offered on data sets, including data not previously permitted to be used or accessed for providing services (telephone log records or mobile geo-location data, for example). And innovation for the PDS and applications marketplace would be a huge new area of development for startups and large companies alike.

**People win.** By collecting, managing, and authorizing access to their own personal data, users will increase their trust and use of digital realms. This empowers people to work together in communities and groups more efficiently and effectively. Users will be able to see themselves reflected, and participate in transactions more directly with vendors.

**Regulators, advocates, and legislators win.** By protecting people with new frameworks that also encourage innovation and new business opportunities, government can give people useful tools to interact with agencies because user's identities are trusted.

# Where is the Comprehensive Online Privacy Framework?

Bil Corry and Andy Steingruebl {bcorry|asteingruebl}@paypal.com  
PayPal Information Risk Management

Position Paper for W3C Workshop on Web Tracking and User Privacy  
April 28 and 29, 2011 – Princeton, NJ

## Summary

Current discussions involving online privacy are primarily in the context of proposed technical controls, e.g. the various Do-Not-Track proposals<sup>1 2</sup>, and Microsoft's Tracking Protection Lists (TPL)<sup>3</sup>. We believe it is premature to discuss technical solutions without having first developed a comprehensive online privacy policy. The history of web cookies should give pause as it provides a compelling example of how a technical solution that precedes policies can have an unfortunate outcome.

We strongly believe that a comprehensive online privacy framework can only be achieved by including all stakeholders, clearly defining ambiguous terminology (e.g. "tracking", "third party"), enumerating user choice and expectations with regard to privacy, developing and testing the use-cases "Do Not Track" will be applied to, carefully considering the impacts and costs of proposed policies, exploring the potential for unintended consequences, then and only then defining the technology we need to enable a comprehensive online privacy framework.

We see the Mozilla proposal for the DNT Header – especially how it avoids over-specifying how the header is interpreted and applied – as the most rational, balanced first step toward a more comprehensive framework. Other proposals such as the Microsoft TPL proposal go too far, too soon and in a very confusing direction for both service providers and users. That said, all the current proposals are putting the cart before the horse.

Technology alone cannot solve the online privacy issues, nor can policy. By carefully crafting the two, a complementary privacy system can be developed.

## Web Cookies and Privacy Failure – Doomed to Repeat?

Although it has been many years, this is not the first time there has been considerable interest in online privacy. David M. Kristol, the original editor of the IETF cookie specification, has written a paper<sup>4</sup> on the history and lessons learned from the cookie specifications, including a substantial amount of history on

---

<sup>1</sup> <http://dnt.mozilla.org/>

<sup>2</sup> <http://datatracker.ietf.org/doc/draft-mayer-do-not-track/>

<sup>3</sup> <http://ie.microsoft.com/testdrive/Browser/TrackingProtectionLists/>

<sup>4</sup> [http://arxiv.org/PS\\_cache/cs/pdf/0105/0105018v1.pdf](http://arxiv.org/PS_cache/cs/pdf/0105/0105018v1.pdf)

the privacy concerns and proposals at that time. He makes it clear that by implementing a technical solution first, without a complimentary framework, it proved too challenging to marry a privacy policy to the technical controls after the fact. He suggested that for future efforts, we should involve the stakeholders, separate the policy from the mechanism, and know that the mechanism alone couldn't solve all the privacy concerns. Our position paper is echoes his sage advice.

## **P3P and Privacy Failure – Doomed to Repeat?**

P3P is a failed privacy mechanism that was designed to solve some of the privacy issues being discussed for Do-Not-Track. While the criticisms are documented<sup>5</sup>, we wanted to point out how it completely fails for one of the use cases we highlight in this position paper. Google uses the following P3P policy for their sites:

```
P3P: CP="This is not a P3P policy! See  
http://www.google.com/support/accounts/bin/answer.py?answer=151657 for more info."
```

Visiting the above URL provides the following explanation:

```
In some situations, the cookies we use to secure and authenticate your Google Account and store your preferences may be served from a different domain than the website you're visiting. For example, if you sign into a Google gadget on iGoogle, your browser may treat these cookies as a third party cookie (even though you are still on a Google site).
```

```
Some browsers require third party cookies to use the P3P protocol to state their privacy practices. However, the P3P protocol was not designed with situations like these in mind. As a result, we've inserted a link into our cookies that directs users to a page where they can learn more about the privacy practices associated with these cookies.
```

Clearly, careful consideration is required of any privacy mechanism by validating the use-cases that will be impacted by the solution.

## **Policy Should Drive Technology**

Over the last few months of 2010 and into early 2011, we have seen a proliferation of new policy and technical controls designed to help users manage their privacy online and prevent “tracking”; collectively under the name “Do Not Track”. Some of these proposals have focused on communicating a user’s privacy choice to a site, while others have focused on technically controlling how web browsers actually interact with websites and with whom they will send and receive data.

We believe that the technical controls have gotten too far ahead of a substantive discussion about a comprehensive online privacy framework, with too many questions still unanswered. What does “tracking” encompass? How should a website behave when a user asserts “do not track me” via a Do-

---

<sup>5</sup> <http://en.wikipedia.org/wiki/P3p#Criticisms>

Not-Track header? How do we resolve conflicts between laws which require collecting/storing information and proposed privacy policies that may require not collecting that information?

Beyond the above, there are numerous additional questions and edge-cases that must be addressed. It is premature for us as a standards-setting community to commit to long-term technical controls for privacy with so many outstanding issues. We already saw how developing mechanisms before policy failed with web cookies. And we're seeing it again with the "Do Not Track" policy discussions being framed by the technical implementations – the technical implementations are in effect forming the de facto standard for "Do Not Track". It is the wrong approach to take if indeed we are concerned about creating a long-term framework for managing online privacy.

## How to Proceed

We believe that in order to make progress in creating comprehensive online privacy standards, the work should proceed in two steps. First, there is much work to be done to define terminology and goals. Second, we believe that any solution proposed must be a complimentary combination of public policy and technical implementations, field-tested against common use-cases collected from a broad cross-section of online service providers.

## Define Terms and Goals

There are many ambiguous terms being used in the Do-Not-Track discussion. At least two of these terms, "tracking" and "third-party", are the most often used, and needing definition.

Just as with the debate about privacy controls in web browsers, there is no generally accepted agreement as to what "tracking" means, who is doing the "tracking", and what data constitutes "tracking".

In the same way, "third-party" is often defined purely in terms of a technical manifestation, i.e. DNS domain names, rather than as is typical in the legal context. Bringing clarity to these terms is critical to making progress in this space.

### *Defining "Tracking"*

So that we can at least have some baseline discussions, we'll settle on the definition that the Center for Democracy & Technology provides in their paper, WHAT DOES "DO NOT TRACK" MEAN?<sup>6</sup>

*Tracking is the collection and correlation of data about the Internet activities of a particular user, computer, or device, over time and across non-commonly branded websites, for any purpose other than fraud prevention or compliance with law enforcement requests.*

Even this definition though may go too far. Depending on the meaning of "non-commonly branded websites", many online "mashups"<sup>7</sup> online must automatically be considered a form of illegitimate tracking.

---

<sup>6</sup> <http://cdt.org/files/pdfs/CDT-DNT-Report.pdf>

Under some policy interpretations of Do-Not-Track:

- A user that uses a mashup of Google Maps and Craigslist Apartment listings must not be logged and tracked in the profile of either of these services, despite making connections to both Google and Craigslist to retrieve data.
- A user of at least one popular flight-pricing website that performs queries via mashup and client-side data aggregation, must not be logged and categorized by any of those airline websites visited even if the user already has a relationship with them.

We believe that fundamentally there must be a broader distinction made between data used in logging transactions, including data not used for fraud and security purposes, and data used to build individual user profiles which enables future behavioral profiling.

Additionally, the scope of “tracking” must be made more distinct so that discussions about what data is to be kept “private” and from whom, is clearer and more universally understood. As things stand today, there is a considerable lack of clarity of whether Do-Not-Track protects a user from:

- Third-Party data aggregators such as online advertising providers
- First party data collection
- Government requests for data

Fundamentally we believe that the obligations on those collecting data should follow the “Use-and-Obligations Framework” developed by the Business Forum for Consumer Privacy.<sup>8</sup>

### *Defining “Third-party”*

The term “third-party” is often used in the technical context when discussing HTTP Cookies to mean a cookie whose “second-level domain” differs from the one the user currently sees in their location bar. Unfortunately, the term “third-party” has an entirely different meaning in other contexts. The basic Wikipedia definition is – “**Third party** is often used to refer to a person or entity who is not one of two involved in some relationship”.<sup>9</sup>

Unfortunately on the web, notions of ownership and contractual status are neither readily apparent nor manifest in the DNS. As such, it is impossible to know merely by looking at a domain name who the owner is and whether it shares an owner with another domain name. Additionally, it is impossible to tell from two domain names what relationship they share contractually with respect to the services they provide and data they collect. Several examples will perhaps help illustrate:

- fb.com and facebook.com are both operated by Facebook but are used for different purposes.
- www.apple.com is operated by Apple Inc., but metrics.apple.com is operated by Adobe’s Omniture group that performs, among other functions, web analytics.

---

<sup>7</sup> [http://en.wikipedia.org/wiki/Mashup\\_\(web\\_application\\_hybrid\)](http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid))

<sup>8</sup> [http://www.huntonfiles.com/files/webupload/CIPL\\_Use\\_and\\_Obligations\\_White\\_Paper.pdf](http://www.huntonfiles.com/files/webupload/CIPL_Use_and_Obligations_White_Paper.pdf)

<sup>9</sup> [http://en.wikipedia.org/wiki/Third\\_party](http://en.wikipedia.org/wiki/Third_party)



- [www.paypal.com](http://www.paypal.com) and [www.paypalobjects.com](http://www.paypalobjects.com) are both operated by PayPal and both domains are integral to the operation of the [www.paypal.com](http://www.paypal.com) website.

These examples show how the domain name of a website is a poor substitute for the purpose and use of that website in a given context.

Any attempt to define “third-party” that does not take these examples into account and instead tries to apply a blanket un-nuanced definition could cause unintended collateral damage.

## Complimentary Policy System

In order to create a comprehensive online privacy framework, it must include a complimentary combination of policy and technical controls. Just as in designing security into software, a threat model and set of requirements must come before technical implementation.

### *Policies and Requirements*

Any discussion on privacy online must start with both well-defined terms and goals. Indeed, a well-stated set of objectives to be achieved by new privacy-enhancing policies and technical controls is a prerequisite to designing those technical controls. As part of defining the objectives, use-cases should be established so as to allow the stakeholders to identify gaps, conflicts or other impacts with their own specific situations.

### *Technical Solution and Controls*

Given our position that much work is still to be done in order to create comprehensive online privacy frameworks, we are unable to at this time make recommendations on what sets of technical controls should be implemented because *we simply don't know*. We are likewise reluctant to consider adopting the current technical implementations as we believe they are premature and prone to harm privacy. Privacy online is nuanced and tricky, as the Federal Trade Commission's 122-page report concludes<sup>10</sup>. Self-evident, if it was as easy as building in some technical controls, we wouldn't be having this discussion.

---

<sup>10</sup> <http://ftc.gov/os/2010/12/101201privacyreport.pdf>

## W3C Workshop on Tracking and User Privacy

### Comments of the Software & Information Industry Association (SIIA)

David LeDuc – March 25, 2011

SIIA is the principal trade association of the software and digital information industry. The more than 500 members of SIIA develop and market software and electronic content for the business, education and consumer markets.<sup>1</sup> SIIA's members are software companies and information service companies, including companies that both provide and rely on Internet advertising. As leaders in the global market for software and information products and services, our membership consists of some of the largest and oldest technology enterprises in the world, as well as many smaller and newer companies.

For over a decade, SIIA has worked with policymakers at the Federal and state levels in the United States, and also with policymakers in Europe, Canada and other regions, to examine the implications and operations of privacy and related laws. This has included work with the relevant Federal agencies implementing existing privacy and security regulations and policies, notably, the Federal Trade Commission's (FTC) approach on unfair and deceptive trade practices, as well as implementation of the Gramm-Leach-Bliley Act (GLBA), the Health Insurance Portability and Accountability Act (HIPAA), and the Health IT Act.

SIIA appreciates ongoing efforts, both within the government and industry groups such as W3C, to assess the important issue regarding online privacy, behavioral advertising and web tracking. SIIA recognizes the critical objective of many policymakers and consumer interest groups to enable individuals with a simple way to opt out of the collection and use of data regarding their online browsing activities.

This has been done for some time, typically through the use of persistent cookies which would signal an individual's choices to various Internet actors. Opting out of an ad network in this fashion does not mean that the user will no longer receive ads. It means that the network will no longer deliver ads based upon the user's web site visits.<sup>2</sup>

---

<sup>1</sup> Our website can be found at: [www.siaa.net](http://www.siaa.net)

<sup>2</sup> See the Network Advertising Initiative's opt out program at [http://www.networkadvertising.org/managing/opt\\_out.asp](http://www.networkadvertising.org/managing/opt_out.asp)

SIIA strongly supports the balance between privacy and the free flow of information, as well as the balance between the need for consumer confidence and continued innovation on the Internet. In an era of rapidly changing technology and business models, SIIA strongly supports a privacy framework that is industry-led, voluntary and enforceable.

With respect to “tracking” and behavioral advertising, SIIA believes that efforts to provide consumers with a clear and easy opt-out should also be effectively focused on addressing potential harm, and they should be mindful not to undermine the benefits of online behavioral advertising. A recent study estimated that targeted ads generated almost three times the revenue of regular run of network ads and accounted for 18% of the total website ad revenue.<sup>3</sup> As many website publishers themselves have noted, restrictions on advertising through ad blocking would risk undermining their economic basis.<sup>4</sup> While SIIA doesn’t oppose creation and use of ad blocking mechanisms, it is critical to recognize the economic harm that could come from confusion between ad blocking and tracing protection—to date there seems to be much confusion in this area, with many of the solutions being marketed as tracking protection that are largely ad blocking devices.

The W3C is a very useful form for assessing the various marketplace options provided and to ensure that private sector mechanisms are effective in promoting consumer choice and preserving the benefits of online behavioral advertising. SIIA is closely studying this issue and we have engaged our members in a dialogue to help private sector development of effective solutions. Following are some of our key conclusions to date:

- **Tracking must be clearly defined** – First, in order to have a productive discussion about online tracking, it is critical that the discussion be focused on the definition of tracking. Generally, “tracking” is not clearly defined. That is, tracking information has numerous potential uses other than targeted online behavioral advertising. Outside of any advertising context, many software and information companies use consumer data to deliver personalized services and to deliver content to users based on information they know about the user, such as improving search and better

---

<sup>3</sup> Network Advertising Initiative, Study Finds Behaviorally-Targeted Ads More Than Twice As Valuable, Twice As Effective As Non-Targeted Online Ads, March 24, 2010 available at [http://www.networkadvertising.org/pdfs/NAI\\_Beales\\_Release.pdf](http://www.networkadvertising.org/pdfs/NAI_Beales_Release.pdf). The study was done by Howard Beales, former director of the FTC’s Bureau of Consumer Protection.

<sup>4</sup> See Ken Fischer, Why Ad Blocking is Devastating the Sites You Love, ArsTechnica, March 6, 2010 at <http://arstechnica.com/business/news/2010/03/why-ad-blocking-is-devastating-to-the-sites-you-love.ars>

tailoring applications and offerings to customers based on their preferences. It is often used for fraud prevention, risk management, control of spam and malware, intrusion prevention or detection.

Additionally, movement within a company's own website or suite of products is clearly not the kind of tracking that consumers are concerned about, and it is vital for businesses to track this kind of movement in order to optimize the performance and appeal of their websites. Similarly, websites routinely log the identity of the websites from which visitors arrive and to which they go when they leave. This provides valuable information about what attracts visitors to the site and what provides them with an incentive to leave. In a voluntary choice regime, these tracking activities would be permitted.<sup>5</sup>

- **Voluntary efforts are best suited to address the goals** – SIIA thinks that a mandated Do Not Track regime or a regulatory requirement to this effect would likely have harmful effects. Government-mandated anti-tracking mechanisms might short circuit the development of these valuable uses of tracking information. On the contrary, voluntary do not track initiatives would likely be better able to accommodate valuable uses while still allowing appropriate user control. SIIA is encouraged by many of the mechanisms under development by industry to inform and provide choice to consumers. SIIA is confident that voluntary choice mechanisms will sufficiently balance the needs of consumers, advertisers and content sites. The voluntary compliance by all Internet actors with the robots.txt protocol is a good example of how a voluntary system can produce desirable policy results without a government mandate.
- **Tracking Protection Lists (TPLs) present many undesirable outcomes** – First, the name is misleading. TPLs are focused not only on enabling users to block just tracking cookies, web beacons and other tools to track movements and activities on the web, but also block ads entirely.<sup>6</sup> The early providers of TPLs include the ad blocking services of EasyChoice, Privacy Protection and Abine.<sup>7</sup> Some users might want to

---

<sup>5</sup> Peter Eckersley makes many of these points in his commentary, "What Does the "Track" in Do Not Track Mean?" February 19, 2011 at <https://www.eff.org/deeplinks/2011/02/what-does-track-do-not-track-mean>

<sup>6</sup> Ed Bott, IE9 and Tracking Protection: Microsoft disrupts the online ad business, February 13, 2011 at <http://www.zdnet.com/blog/bott/ie9-and-tracking-protection-microsoft-disrupts-the-online-ad-business/3004>

<sup>7</sup> Ed Bott, Privacy protection and IE9: who can you trust? February 14, 2011 at <http://www.zdnet.com/blog/bott/privacy-protection-and-ie9-who-can-you-trust/3014?pg=2>

block ads in addition to blocking the tracking cookies, web beacons, and other devices that can track their movements from web site to web site. But it is not useful, and potentially harmful to the economic model of the Internet, for ad-blockers to be mislabeled as tracking protection solutions. There are technological solutions that can effectively prevent tracking and still allow for ad placement that isn't behaviorally targeted. It is these solutions and only these solutions that should be described as tracking protection. It is critical that discussions are clear on this point.

Further, there is a substantial possibility of consumer confusion regarding these lists. TPLs do not easily reveal the parties/domains blocked, so users may think they are blocking only bad actors, but in fact end up blocking sites that they actually want to see. In order for this to be implemented effectively, there is a high level of technical understanding necessary by the user. Even in such cases, accidental blocking is quite likely. Also, some TPLs allow third-party domains listed to be displayed. Other TPLs are exclusively blocking lists. When a user installs multiple lists, hierarchy rules provide that "allow instructions" trump "block instructions." This is inherently confusing to users and can create big problems.

- **From a technological and user experience perspective, SIIA would like to see continued support for voluntary opt-outs of tracking** – Persistent opt-out cookie initiatives have proven to be a highly effective mechanism for easy opt-out. Implemented as either a plug-in or a native component of browsers, this approach can provide a highly effective way for users to opt out of personalized advertising from participating networks and store the setting permanently. Importantly, the focus on the technological activity of "tracking," such as managing cookie controls, seems to be a highly effective approach, more so than approaches that simply shut off crucial parts of web pages and ultimately threaten to compromise user web experience.
- **The focus on browser web tracking is quite limited** – Less and less Internet activity is conducted through the browser and more is being done through applications such as instant messaging, voice over internet, RSS feeds, and streaming video. These applications use the Internet's underlying communications protocols, but they do not

use the browser capabilities.<sup>8</sup> By virtually all accounts, these trends represent the future of the Internet. Therefore, while a browser-based do not track mechanism is a useful endeavor, it is generally a narrow approach to the greater challenge of providing users with choice on “tracking.” Again, in this broader effort, a voluntary initiative would be best suited to handle technological innovations and developments of this nature, and the W3C is a well-suited forum to discuss among key stakeholders.

---

<sup>8</sup> Chris Anderson and Michael Wolff, “The Web is Dead. Long Live the Internet,” Wired, August 17, 2010 [http://www.wired.com/magazine/2010/08/ff\\_webrip/all/1](http://www.wired.com/magazine/2010/08/ff_webrip/all/1)

# Tracking Transparency

Wendy Seltzer\*

Berkman Center for Internet & Society at Harvard University  
and Princeton Center for Information Technology Policy  
wendy@seltzer.org

March 31, 2011

The Fair Information Practice Principles<sup>1</sup> model of privacy protection, influential in U.S. privacy law, depends upon transparency. Transparency is key to adequate notice, which in turn is necessary to meaningful choice, access, and enforcement. Thus transparency to end-users is a critical component of any do-not-track mechanism. I compare the transparency of server-side header response and client-side request-blocking and suggest that the latter is more directly transparent in its operation.

Tracking itself poses transparency challenges: trackers know more about their prey than is apparent to the typical Internet user. Much tracking happens through back-end correlation, building up server-side profiles outside the view of the user; even what is sent to the browser is often hidden under browser rendering (or non-rendering, in the case of third-party cookies and transparent or zero-pixel images); and users are at information disadvantage to their trackers. Moreover, wariness of the “creepiness” factor, along with simple scale economies, may cause trackers to tune the profiling less finely than their data would make possible. Users are therefore rarely exposed to a direct mirroring of all data collected from them, or the full customization that profiling would make possible. Even as their experience is being customized, users generally have no window into others’ experiences for comparison. While advertisers speak publicly of discounts to good customers, the customers worry about price discrimination that charges more to those with demonstrated willingness to pay.

Technological tracking protection measures can give end-users greater control of their privacy choices – but their effectiveness will depend to a great extent on the notice they give users of what is being defended against, and how. While the current Firefox 4 and IE 9 both employ a new HTTP header, DNT, they implement it differently, changing the level of user visibility and control.

In Mozilla’s Firefox 4, users navigating to the Advanced, General tab are shown the option “Tell websites I do not want to be tracked.” As the online help indicates, this option relies entirely on the recipient for

---

\*Affiliations listed for identification purposes only. Comments reflect personal position, not that of any institution.

<sup>1</sup>FTC Fair Information Practice Principles <http://www.ftc.gov/reports/privacy3/fairinfo.shtm>.

implementation: “Checking this box will tell websites that you wish to opt-out of tracking by advertisers and other third-parties. Honoring this setting is voluntary – individual websites are not required to respect it.”<sup>2</sup> Before this setting can have any impact, websites and their tracking partners must learn to recognize the DNT header and determine a policy response to it.

The DNT policy decisions start with specifying what “tracking” means: Is it following a user across multiple sites or multiple sessions, or does it include watching repeat intra-session visits to the same website?<sup>3</sup> Is it correlating browsing behavior with personally identifying information gained from user input or environment? Could a website assert that it was not “tracking” if it merely collected the information but didn’t *use* it? Unless all sites respond using the same definitions, the header will operate differently from site to site.

Thus although the DNT header in Firefox’s user-option is relatively simple to exercise, its operation is not very transparent. A user trying to determine the real impact of that toggle would have to visit all of the websites of concern, seek out privacy policies, hope those policies also bound third-party trackers or watch page-loads and requests to ferret out all these “partners” and seek out their policies as well. Even after this sleuthing, the user would still have to rely on the sites to describe and abide by their stated policies, or depend on enforcement authorities such as the U.S. Federal Trade Commission to police against “unfair or deceptive acts or practices.” Users will get little or no feedback to let them know whether their do-not-track preferences are being honored on a site-by-site basis.

Microsoft’s IE 9 instead offers users a configurable browser-side block-list to control what gets sent to online trackers: “Tracking Protection Lists are like ‘Do Not Call’ lists for third-party content on a website. By adding a Tracking Protection List or ‘TPL,’ you can control whether your information is sent to third parties listed on the TPL.”<sup>4</sup>

If the user configures this list by choosing from among the third-party TPLs Microsoft links to, he or she can review the contents of the list. These require some investigation: in particular, if the user installs the current TRUSTe list, he should note that “TRUSTe’s TRUSTed Tracking Protection List *enables* relevant and targeted ads from companies that demonstrate respectful consumer privacy practices and comply with TRUSTe’s high standards and direct oversight.” (emphasis added)<sup>5</sup> In fact, the current *easy.tpl* *allows* tracking from 1570 domains, among them *freecreditscoreonus.com*, *cashadvance.com*, *LipitorLink.com*, *makingsenseof-painrelief.org*, and *acxiomdigital.com* while blocking only 23 domains. TRUSTe’s own site promotes this list “to block companies that offer poor privacy protection, while ensuring that trustworthy companies who

---

<sup>2</sup>Firefox Help, <http://support.mozilla.com/en-US/kb/Optionswindow-Advancedpanel>

<sup>3</sup>In the age of suspend, how many users keep single sessions open for weeks?

<sup>4</sup>Microsoft, Tracking Protection Lists, <http://ie.microsoft.com/testdrive/Browser/TrackingProtectionLists/faq.html>

<sup>5</sup><http://ie.microsoft.com/testdrive/Browser/TrackingProtectionLists/>



protect their privacy can continue to provide them with a richer, more personalized browsing experience.”<sup>6</sup> Following the link to TRUSTe’s “Third Party Data Collection Certification Program Principles,” however, one finds that all the TPL buys is protection against the linking of personally identifying information and the option to read another privacy policy and opt out *again* if one wants not to be tracked on any of these 1570 domains.<sup>7</sup> Other lists Microsoft links, from Abine, Easy List, and Privacy Choice, appear to be more straightforward block-lists, but as the spec is designed, if a user chooses multiple lists, the “ALLOW” from any list takes precedence over “BLOCK.”

While some of its current implementations may be surprising, however, the operation of the Microsoft-implemented tracking-protection is transparent to the end-user in two ways – its lists are user-side, and some of their effects are directly visible in the browser: a blocked ad does not show up, often leaving a blank space. This mechanism can be bolstered by self-regulatory programs or regulatory policing, but it does not depend on them.

Users can learn to protect privacy – and to determine which aspects of tracking are too invasive – if they get feedback on how their choices change their experiences. If choice is to be meaningful, and if users are expected to understand tracking as part of a bargain for online content, we need to assure that they have a real-time view of what they are exchanging.

I look forward to participating in the W3C workshop to explore how tracking protection technologies can help give end-users visibility into and control of their online experiences.

---

<sup>6</sup><http://tracking-protection.truste.com/>

<sup>7</sup>Cashadvance.com, for example, discloses in its linked privacy policy that “All cookies served on the domain, both session and persistent are tied to the Personally Identifiable Information you provide,” and that “This privacy statement applies solely to information collected by Company even in cases where we may frame another site with our own.” <http://www.cashadvance.com/privacy-policy> It proudly displays a TRUSTe seal and appears with a “+d [ALLOW]” on TRUSTe’s TPL.

## Security and Fraud Exceptions Under Do Not Track

Christopher Soghoian  
Center for Applied Cybersecurity Research, Indiana University

Position Paper for W3C Workshop on Web Tracking and User Privacy  
28/29 April 2011, Princeton, NJ, USA

### Introduction

As the debate over Do Not Track continues to evolve, the most important issue under discussion is the definition of tracking. Generally speaking, advertising networks and other firms that engage in online tracking wish for this definition to be as narrow as possible, while privacy advocates are pushing for a broad definition with few exceptions.

Even among many privacy advocates, there seems to be a general acceptance that companies should be able to engage in some forms of tracking and data collection in order to protect against fraud and security related threats.<sup>1</sup>

The fraud and security issues are particularly challenging, because many companies are unwilling to publicly disclose how much data they need, or how they use it, for fear of tipping off those who would misuse the information. As a result, we are forced to take these companies at their word, without the means to independently verify that they do in fact legitimately need the information they are tracking, and that they need to retain it for as long as they are doing so.

Unfortunately, this exception has the very real potential to swallow the rule. For example, in 2008, primarily in response to strong pressure from European privacy regulators, Yahoo! announced a bold new policy of only retaining identifiable log data for search and other services for 90 days. However, the company keeps a second set of identifiable logs for six months, which it uses for fraud and security related purposes. Although Yahoo! will not reveal how much data is kept in this alternate set of logs, these files can of course be obtained by law enforcement agencies wishing to learn how users interact with Yahoo!'s site, long after the primary database of logs have been anonymized.

There are of course different privacy concerns related to Yahoo!'s first party collection of search query information and data collected by third party advertising networks. While law enforcement agencies have shown a keen interest in search queries, I am not aware of any advertising network that has received queries from law enforcement agencies for information about users' web browsing activities. Nevertheless, the Yahoo! example does serve to demonstrate that data retained for security and fraud purposes can seriously undermine the effectiveness of an

---

<sup>1</sup>The IETF draft proposal for Do Not Track by Mayer *et. al.* includes security and fraud related exceptions to the definition of tracking. Likewise, the DNT scoping proposal published by the Center for Democracy and Technology includes an exception for "Data collection required by law and for legitimate fraud prevention purposes."

otherwise privacy-preserving data retention policy, particularly when companies are unwilling to reveal what data is being retained and how long they are keeping it.

### **First party activities are not considered “tracking”**

When the average consumer, regulator or policy maker is told that some kinds of tracking are necessary for purposes of fraud and security, the argument of course sounds reasonable. No one wants to allow fraud or hacking, particularly given that consumers ultimately pay the cost, for example, in the form of higher credit card interest rates and transaction fees.

However, when you actually enumerate the most common examples of tracking for fraud prevention, it quickly becomes clear that most of them do not fall under the any of the definitions of tracking under consideration, even without the fraud and security exceptions. Consider the following scenarios:

- A consumer logging into their online bank account and then paying a bill.
- A consumer clicking on a Facebook “Like” button while visiting a blog.
- A consumer conducting a Google search, and then clicking on one of the search results.
- A consumer clicking from a merchant’s shopping cart to Paypal, where they authenticate and then pay for a product.
- A consumer purchasing an item at Amazon or Walmart’s online store.

In all of these scenarios, the consumer is interacting with a website in a first party manner. No one is suggesting that PayPal should not be able to track a user when they visit their site, that Facebook should not be able to log the clicking of Like buttons to protect against click-jacking,<sup>2</sup> or that Bank of America should not be able to use first party Flash cookies as part of its SiteKey two-factor authentication system.

Next, consider the legitimate desire (and obligation) of third party ad networks to protect against click fraud, in which a malicious first party publisher generates fraudulent clicks for ads displayed on its own site. These ad clicks are a first party activity (or should be considered as such), because the moment a user clicks on the ad network’s banner advertisement, the user is knowingly interacting with that company’s servers. Certainly, as soon as the click is processed, the user will leave the publisher’s website and be taken to the website of the advertiser, who can now drop cookies into the user’s browser as a first party.<sup>3</sup>

---

<sup>2</sup>Click-jacking issues aside, Facebook logging the clicking of the Like button seems to be a clear first party interaction. However, Facebook logging the display of the Like button before it is clicked is almost certainly a third party interaction, and should be prohibited when the user has enabled Do Not Track.

<sup>3</sup>There are of course advertisements that a consumer can interact with without leaving the publisher’s site. As such, there edge cases that are worthy of further discussion. An example raised by Ashkan Soltani is that clicking the mute button on an auto-playing ad should not be considered a first party interaction.

By excluding these legitimate activities from the definition of tracking, only a few third party forms of tracking remain for us to consider.

### **Tracking for the purpose of detecting advertising impression fraud**

Many advertising networks deliver advertisements under a pay for impression (CPM) model. In order to bill their clients, the advertisers, they need to be able to demonstrate that the 1000 ad impressions were delivered to 1000 different users, and not the same user clicking the reload button 1000 times. This is currently done by giving users unique tracking cookies, and logging impressions.

Before attempting to evaluate the tracking activities necessary to combat ad impression fraud, two important factors should first be considered:

- Apple's Safari browser has long blocked the setting of third party cookies by default. Even so, ad networks still monetize the impressions generated by the millions of consumers using Apple's products.
- An adversary seeking to engage in impression fraud can always delete, modify or refuse to accept cookies. As such, ad networks cannot trust cookies sent to them by adversaries.

These two factors mean that many advertising networks already detect and prevent impression fraud without the benefit of cookies or other unique identifiers.

It would seem rather illogical to permit ad networks to continue to use unique cookies to track users who have expressed a strong desire to not be tracked, when these ad networks already have to make do without giving cookies to millions of Safari users who have expressed no privacy preference at all. As such, I think there is a strong argument to be made that advertising networks should be prohibited from tracking users via cookies or other locally stored unique identifiers when a user has expressed a desire to not be tracked (this could be enforced via legislation, or preferably, by the browsers refusing third party cookies or at least making them session only).

With regards to logs kept by ad networks, the sensitive information is not really the user's IP address, but the information contained in the referring header revealing the first party site that the user was visiting when the advertisement was displayed. In some cases, the privacy concerns could be addressed by redacting the path portion of the URL (webmd.com vs webmd.com/cancer/). However, in other cases, the domain name itself would be sufficient to raise privacy concerns (for example, a gay dating website, or a website focused on a specific medical disease). Because some URLs raise greater privacy issues than others, the only automated way to protect this information reliably would be to redact the entire referring URL.

However, it is likely that advertisers wish to know, even with just aggregate numbers, which specific URLs are generating the most impressions (and clicks) in their advertising campaigns.

As such, the most practical solution to protecting privacy with regard to impression log data may be to use a combination of front end data anonymization (for example, hashing IP addresses) and relatively short retention times.

### **Tracking for security purposes**

Third parties, like first party sites, have a legitimate interest in protecting the security of their systems. This includes detecting and protecting against denial of service attacks and intrusions.

In the case of denial of service attacks, logging can be used to detect large numbers of requests from the same IP address, although this is less useful when the attack is distributed among a large pool of IP addresses. It is unclear though why long data retention periods are necessary to protect against such attacks. Furthermore, if a particular IP address is not generating traffic above some reasonable threshold, it is not even clear why logs are necessary at all.

In order to protect against intrusions and other sophisticated attacks, companies obviously want to know how a potential attacker is interacting with their servers. Of course, hackers do not identify themselves as intruders beforehand, and so sites must log every single request in order to later determine which particular requests were associated with a hacking attempt.

While it would be unwise to try and dictate what data third parties can and should collect in order to protect their systems against skilled attackers, it is worth noting that all companies face the problem of security breaches and denial of service attacks. As such, there isn't likely to be any particular "secret sauce" specific to protecting third party sites from attack. Unlike sector-specific attacks such as click fraud and ad impression fraud, it should be possible to have a relatively open discussion about the data retention and tracking necessary to reasonably protect against the general security threats faced by all firms.

### **A word on fingerprinting**

The use of browser fingerprinting presents a unique problem to those concerned about user privacy. First, users do not know when their browsers are being fingerprinted, and second, users often are not given a way to opt out, at least when fingerprinting is used for fraud prevention.

As a baseline requirement, fingerprinting should be disclosed, when conducted by first or third parties. Not only can consumers not easily determine that a site is collecting a fingerprint of their browser, but few companies will confirm their own use of these technologies, even when directly queried by privacy advocates.<sup>4</sup>

---

<sup>4</sup>Employees at one prominent first party company would not comment on their own use of fingerprinting technology when I asked. Such silence is disgraceful, and suggests that these companies know they are engaged in a practice that would cause outrage among consumers and legislators if disclosed.

If first parties wish to fingerprint browsers, they should be required to clearly and prominently notify users that it is occurring. This does not mean the website needs to reveal which specific data points are collected and analyzed, but simply that the website is collecting information about the user's browser that will be used to identify them the next time they visit.

Third parties should be prohibited from using fingerprinting technology, preferably at all times, and at least when a user has enabled a Do Not Track setting in their browser. While there may be legitimate scenarios in which this third party collected information may benefit first parties who wish to protect themselves from fraud, the covert collection of data by these third parties raises far too many privacy issues. Third party fingerprinting is still new enough that it can be quietly killed off without seriously disrupting the market. Now is the time to do, before large numbers of first parties become dependent upon this highly problematic source of tracking data.

### **Conclusion**

The development of Do Not Track policies and technologies promise to deliver a significant increase in privacy protection for the average user. Of course, as the industry continues to remind us, there are some legitimate forms of tracking, and some of these relate to the prevention of fraud and protection of site security.

As is also the case in the area of national security, there is a great risk that those wishing to abuse their powers may hide their otherwise improper behavior behind the shroud of "security."

Technologists and regulators should be highly skeptical regarding companies' claims of security and fraud, at least when they are unwilling to reveal the exact data they need to track, and how long they wish to keep it. Many such claims cannot, and will not stand up to reasonable analysis.

## **TRUSTe Position Paper for W3C Workshop on Web Tracking and User Privacy**

TRUSTe has been actively involved in privacy compliance programs for websites, 3<sup>rd</sup> party ad and data providers, applications and cloud services. The recent upswing for enhanced privacy programs regarding 3<sup>rd</sup> party tracking and online behavioral advertising has led to a mix of self-regulatory programs and new consumer features added to recent browsers such as IE9 and FF4.

TRUSTe has deployed a mix of solutions ranging from a DAA-approved Notice and Choice program to certification programs for data companies plus technology for consumers to utilize for preference management. TRUSTe has been providing services for consumers that help them identify companies that meet a minimum industry best standards bar to signal good privacy and data governance practices. TRUSTe has also offered objective preference services that enable consumers to select from a range of participation options that meet their personal preferences and values.

For many consumers, TRUSTe's brand has served as short cut to understanding this very complex calculation of what good privacy means. TRUSTe has provided these services in ways that participate in self-regulatory regimes and those that specifically offer Safe Harbors in exchange for meeting minimum standards as mandated by governmental organizations.

As the dialog has shifted towards the discussion of a Do-Not-Track header, TRUSTe is interested in participating in the dialog and sharing its experiences in the implementation of these types of solutions, and to help shape the ultimate definition and direction of how technology can be best applied to consumers in what many consider a very nuanced and non-primary part of their online experience.

From a very high level, there are three primary constituencies that form the inflection points of the spectrum to address with respect to privacy controls: 1) Those very sensitive to their privacy and very proactive to learn and use the technical controls available to them to control their experience, 2) Those that either do not understand or do not care about privacy controls and thus do not want overly complex privacy controls, and 3) The large band in the middle which are people that do care to some degree when asked, but do not take the time to proactively manage any sort of privacy controls unless some incident has happened to them personally.

Any sort of controls that are offered directly to consumers need to consider these constituencies. As has been proven by past controls, if too complex, they will not be used properly. Or like the experience of AdBlock+ showed in the Firefox community, a security perspective will prevail, which leads to locking down everything and taking the most conservative path forward with respect to privacy. However, controls need to have the features that address those users that want the deep granular control.

## TRUSTe Position Paper for W3C Workshop on Web Tracking and User Privacy

With respect to the ad technology layers, most consumers do not understand the various entities in the ecosystem, as these are not companies with consumer-familiar brands. Requiring consumers to assess each is an inappropriate task with respect to its desired purpose. Simplicity needs to be a starting point unless there are easy ways to communicate differentiation among various ad and data companies that might be useful to consumers. Oversight and certification offer options to do this.

From the business or server side, the current defacto choice system has been based upon cookies for opt-ing out of seeing targeted ads, which were historically delivered via a website or industry groups' privacy policy and most recently, moving into the ad unit and on the same page as the ads appear. Ad companies using OBA are required via their self-regulatory organizations to deploy such system by approximately mid-year 2011. There are well-documented limitations to a cookie-based system around persistence and usage-only controls, which has led to browser extensions providing various features including script-blocking that provides control of cookie-based tracking by blocking the scripts that deliver those cookies.

The domain blocking systems, historically AdBlock+ (an all block solution) and MSFT's IE9 Tracking Protection Lists (a combination white and black list approach) presents a solution where companies need to consider their position on the most distributed lists to understand whether consumers are seeing their ads or not, as these solutions block the full ad content in addition to the collection mechanisms.

TRUSTe has built a TPL to work with the IE9 program which will provide a balanced list of ALLOWed and BLOCKed companies offering consumers a choice to see relevant ads, but only from companies that respect privacy per TRUSTe's documented standards. The qualifications for the ALLOW require a certification program that elevates only the best companies and requires DAA deployment where applicable.

With respect to the DNT header, there are both technical and political paths in consideration, of which the latter is out of scope of this position paper. A DNT header presents a more simplified preference for users to indicate this preference once and universally. However, this implies that the consumer understands what "tracking" is and what the implications of selecting a universal "opt-out" is.

From the business side, there is the open set of questions of compliance: If a consumer signals their preference how do they know a particular ad server (1) received it and (2) honored it? Assuming there is a methodology to convey this acceptance and honoring, how will this information be managed in the event of a dispute? How will information about the consumer's preferences be managed to determine if companies did not receive or honor it, and will this change the user experience?



## **TRUSTe Position Paper for W3C Workshop on Web Tracking and User Privacy**

Additionally, how can companies that provide the necessary industry requirements be recognized positively versus other companies that just ignore this system?

These and other questions deserve careful consideration; else they can relegate this technology to a partially adopted and confusing state that would reduce its effectiveness.

TRUSTe can contribute by helping companies deploy the necessary program elements to get into compliance with this system and by providing consumer-friendly approaches to demonstrate their respect of the opt-out preference to elevate their good standing and present the user with multiple preference options.

W3C workshop on web tracking and user privacy  
Deutsche Telekom AG, Group Privacy

Position paper

#### a) Background

Web tracking and data privacy are two closely related issues for Deutsche Telekom. Mechanisms enabling acquisition and use of context-based information for enhancing our diverse connected life and work offerings are becoming increasingly significant. We aim to bring transparency into this process of acquiring and utilizing user-specific information and open it up to influence by the user within the scope of the law.

Individual and user-specific profiles that enable personal addressing of users, e.g. by e-mail, can only be created if the user in question opts in to the process.

In addition, current German data privacy law permits the creation of usage profiles under pseudonyms for online content offerings. This situation must be made clear to users and they must be given the option of opting out of profile creation.

On the technology side, such creation of usage profiles under pseudonyms is carried out using cookies. The German data privacy authorities have published a guide to teleservices and media services which recommends the use of opt-out cookies to enable users to object to profile creation.

At European Union level, the July 2002 cookie directive was updated in November 2009. The updated version must be implemented in national law by the end of May 2011 and brings with it changes to several aspects of the previous cookie directive.

The revised version of Article 5(3) 2009/136/EC states that website operators can only use cookies in future if the user gives his or her consent, having been provided with clear and comprehensive information. The directive does not specify how such consent can be obtained. It was, however, mentioned in comments on the directive that this legal requirement could also be implemented using the relevant settings options of browsers.

## b) Known problems

With respect to the data privacy law situation, Directive 2009/136/EC is yet to be implemented in national law. There is uncertainty as to which mechanisms or processes to use in order to implement the requirements.

Current legal provisions for data privacy in Germany are relatively new in comparison to other laws, but it is doubtful that they take sufficient account of technical and social developments.

Technical implementation of the opt-out function using relevant cookies, which is required for creating usage profiles with pseudonyms under German law, reaches its limits when users delete their cookies.

In addition, users are forced to indicate their desire to opt out for each portal; from a user point of view, a comprehensive opt-out can only be achieved if users make changes to the general settings that determine how their browser handles cookies, for example, through deletion of all cookies at the end of a session.

In addition, ad server operators carry out separate tracking of users. Cookies are also used for this process. In the vast majority of cases, the function of these cookies is not made clear to portal managers or users.

## C) Expectations and requirements

From a data privacy perspective, preference is to be given to solutions that meet the following requirements:

### Legal conformity

The current requirements of German data privacy law, opt-out for tracking / profiling under pseudonyms and the requirements of EU Directive 2009/136/EC must be comprehensively supported by a technical solution.

### Transparency

Users must be given clear indication of the type of profile created and how it is used. Settings and opt-out options must be presented as clearly as the specific potential benefit to users of creating and using profiles.

### Ease of use

There must be one point where users can influence and, if relevant, manage the tracking of their usage data and the creation of profiles.

#### Deleting data

Users must be given the option of deleting the profile data assigned to them.

HYPERLINK "http://www.sachsen-anhalt.de/fileadmin/Elementbibliothek/Bibliothek\_Politik\_und\_Verwaltung/Bibliothek\_LFD/PDF/binary/Service/orientierungshilfen/oh-tele-medien-dienste.pdf" [http://www.sachsen-anhalt.de/fileadmin/Elementbibliothek/Bibliothek\\_Politik\\_und\\_Verwaltung/Bibliothek\\_LFD/PDF/binary/Service/orientierungshilfen/oh-tele-medien-dienste.pdf](http://www.sachsen-anhalt.de/fileadmin/Elementbibliothek/Bibliothek_Politik_und_Verwaltung/Bibliothek_LFD/PDF/binary/Service/orientierungshilfen/oh-tele-medien-dienste.pdf)

# Proposal for a “Down-the-Chain” Notification Requirement in Online Behavioral Advertising Research and Development

David Thaw  
Department of Computer  
Science  
University of Maryland College  
Park  
dbthaw@cs.umd.edu

Neha Gupta  
Department of Computer  
Science  
University of Maryland College  
Park  
neha@cs.umd.edu

Ashok Agrawala  
Department of Computer  
Science  
University of Maryland College  
Park  
agrawala@cs.umd.edu

## 1. INTRODUCTION

Contextual online advertising, also known as behaviorally-targeted or “Online Behavioral Advertising,” is one of the fastest growing markets on the Internet [4, 1, 6]. These terms are used to describe a variety of Internet advertising services, many of which collect information about individuals’ identity, personal characteristics, preferences, and online behaviors. Marketers consider such information quite valuable and use it to deliver advertising content on an individual basis. By customizing the delivery of advertising content, marketers argue that consumers receive information more relevant to their individual needs and wants [7].

Much of the information collected, however, may be of a sensitive nature and/or may conflict with consumers’ privacy expectations [13]. There also is substantial confusion among consumers as to what information is collected and how that information is used. As noted by many scientists [14] and policymakers [2], privacy policies (in their current form) are ineffective at informing consumers of what information collection occurs, what options are available regarding such collection, and how consumers would go about exercising those options.

In this paper, we propose a system called “down-the-chain” notification, under which producers at each step of the research, design, implementation, and maintenance stages bear the responsibility to document the information input “needs” of their algorithms and other technical elements to ensure accurate information is available as to the actual needs of the technology. We feel that such a requirement will help improve consumer options and help ensure those choices are enforceable.

## 2. CONTRIBUTION

As computer scientists and attorneys working in the behavioral advertising space, we feel that informed consumer choice is essential to the continued viability and vibrancy of the Online Behavioral Advertising (OBA) industry. We believe this workshop is an important opportunity for technical policy and business stakeholders to interact. We seek feedback on our proposal and we hope to use this workshop as an opportunity to engage the input of these stakeholders to improve our proposal.

For consumers to be informed effectively, the drafters of consumer notices must have access to complete and accurate descriptions of what information is being collected about individuals and how that information is being used. Likewise, in making decisions regarding the design and implementation of systems, business and technical staff must have access to complete and accurate information describing the data needs of various algorithms and other technical components upon which the systems they implement are based.

This flow of information from the whiteboards where algorithms are first conceived to the end-user/consumer via privacy policies and other notice-and-choice mechanisms is essential to ensuring that:

- The “administrators” of Online Behavioral Advertising systems (e.g., ad networks wishing to collect data and perform analytics) have the maximum number of consumer (privacy) choices available to offer consumer users of the content (e.g., website visitors); and
- The published consumer expectations (e.g., via privacy policies) are as accurate as possible and not erroneous as a result of disconnects between the attorneys and privacy professionals drafting notices and the technical developers actually implementing the design decisions in software.

As experienced computer scientists, we are aware of the challenges inherent in pressing for any standard requiring developers to document their code. However, given the heightened privacy concerns inherent in this space, and the notable gap in technical understanding between those individuals drafting consumer-facing materials and those individuals designing/maintaining the systems, we believe that in

consideration with the potential sensitivity of the information at issue, a higher standard is in order.

Our paper discusses these issues in greater depth, using as an example research currently being conducted at Maryland. We propose a framework for this "down-the-chain" notification and raise several issues for discussion that we feel are still outstanding in our position/proposal.

### 3. DOWN-THE-CHAIN NOTIFICATION

We propose that those who design and implement the technologies enabling OBA have a responsibility to document the "information requirements" of their technologies. Adopting a well-known principle in software engineering called precondition/postcondition documenting, designers and implementers of OBA systems would be required to specify what types of information must be collected and what types of information must be persistently stored for each function of their system to operate.

Since there are different types of (often sensitive) information involved in the OBA system, we suggest a "down-the-chain" notification system where entities involved in any of the five roles as mentioned in Section 3.1, no matter whether they function independently or collaboratively, communicate the information needs of their work to the next role in the chain. So, for example, when computer scientists propose new algorithms they must also state the requirements those algorithms have with regard to what information must be collected and what information must be persistently stored for the algorithm to function. This helps designers building the actual systems make better-informed decisions about what information to keep and what can be discarded (or not collected at all). Moving down the chain, the work of the business professionals and privacy professions becomes easier due to this efficient system of documenting the data requirements at each step.

The goal of this requirement is two-fold. First, to enable choice by allowing decision-makers (business professionals and consumers) to choose not to collect/retain or not to supply any more information than necessary for operation of the system. Second, to help enforce choices by providing visibility into exactly what types of information are used. We explore how this requirement might function in the context of a system currently under development at the MIND Lab at the University of Maryland [3].

#### 3.1 Roles

There are multiple entities or players involved in the design, implementation and usage of OBA system. We discuss this notification requirement in the context of five generalized roles: 1) computer (and other technical) scientists; 2) software engineers and system administrators; 3) business professionals; 4) attorneys/privacy professionals; and 5) end-users/consumers. We define these roles as follows:

- **Computer (and other Technical) Scientists:** Those who design the fundamental algorithms and other technical elements upon which OBA systems are based.
- **Software Engineers and System Administrators:** Those who design and maintain the information sys-

tems that support Online Behavioral Advertising.

- **Business Professionals:** Those who make business decisions according to the requirements of OBA systems.
- **Attorneys/privacy professionals:** Those who draft consumer communications and vet the requirements of OBA systems and software for legal and regulatory compliance.
- **End-Users/Consumers:** The *users* of OBA systems and the websites supported by OBA advertising revenue. These individuals should be able to make informed choices as to their participation in and use of OBA systems and the websites and other technologies OBA revenue supports.

In proposing a down-the-chain notification requirement, our goal is to ensure two conditions:

- The maximum number of choices are made available to each decision-maker in the chain; and
- The choices made are both given effect and are verifiable (auditable).

#### 3.2 Types of Information

In certain cases, for example the protection of proprietary trade secrets regarding system design, it may be desirable to allow developers to report what information must be collected/stored in categories rather than precise (individual) elements. This approach also allows for flexibility in upgrades/modifications if a new data field is added (that does not fundamentally change the privacy landscape). Additionally, this approach may serve to simplify the presentation of the information collection/retention specifications.

To give this approach effect, we propose segregating data elements into the following categories:

##### 3.2.1 Static Demographic Information

Static Demographic information is the demographic information such as gender, age, marital status, family size, user-defined interests, race/ethnic origin and the genetic makeup, religion, etc. This static demographic information does not change in the short term or with user "browsing behavior".

##### 3.2.2 Personal Information

Information explicitly capable of uniquely or near-uniquely identifying an individual:<sup>1</sup>

- Uniquely identifiable information: includes full name<sup>2</sup>,

<sup>1</sup>The categories that follow list example data elements and are not intended to convey comprehensive lists.

<sup>2</sup>Assume full name collisions are resolvable.

SSN, mobile phone numbers<sup>3</sup>, financial account information<sup>4</sup>.

- Near-uniquely identifiable information: includes landline phone numbers, email addresses, mailing addresses.

### 3.2.3 Behavioral Information

Any type of past browsing actions including user engagements such as mouse overs, non-navigation clicks, etc. along with the search queries issued by the user is called behavioral information. Additionally, if knowable, the amount of time spent in any of these activities is called behavioral data.

### 3.2.4 Modeled Information

Modeled Information consists of generalizations about interests, behaviors, etc. based on past behavioral events but without including any specific behavioral events. Modeled information usually includes predictions and rules made by humans or inferencing algorithms over the behavioral and static demographic data.

## 3.3 Inferencing Algorithms

Inferencing algorithms are used to make sense of these vast amounts of data or information collected in OBA systems. Different algorithmic techniques use different feature sets of the data and also have different requirements on the storage time of the data. We can broadly classify the learning algorithms into two categories: offline learning and online learning.

- **Offline Learning** : The traditional method for doing inferencing where models and rules are developed and updated from (static) historical data and then are applied to future data.
- **Online Learning**: A dynamic method of updating the learnt models in real time as new data is encountered.<sup>5</sup>

## 4. ARCHITECTURE OF AN EXAMPLE ONLINE ADVERTISING SYSTEM

To illustrate our “down-the-chain” system, we present here an architecture which will enable developers, designers and system architects to maintain a coherent view of the OBA system as shown in Fig. 1. This example architecture demonstrates how data can be segregated to restrict access to individuals’ personal information while still enabling certain personalization features of OBA.

It is generally preferable from a software engineering perspective to implement privacy features while the system is being designed rather than attempting to “layer” those features on to an existing system[5, 12]. To make informed

<sup>3</sup>Mobile phones may have many-to-one numbers-to-individuals mapping, however this is much less frequent than with email addresses and landline phone and thus we place mobile phones in the uniquely-identifiable category.

<sup>4</sup>Treat joint accounts as single person for the purposes of discussion.

<sup>5</sup>There are also batch-learning algorithms which lie in between of online and offline learning.

design decisions, software engineers must be able to identify what information is required, when the information is required and what elements of the system require the information.

To ensure that the above principles are followed, we divide the system into 3 parts:

- **Globally Unique Identifier (GUID) Table and Personal Information(PI)** : GUID table contains a mapping of username to globally unique identifier which can, for example, be stored in a browser cookie. All other databases are keyed to the GUID. PI is the database containing personal information such as username, full name, passwords, etc. The access to personal information is restricted and “firewalled” off from the rest of the system. PI is a write-only database to prevent behavioral and other data from being linked to individuals’ explicit identity. We consider the PI database to be write-only in the context of our example system for the purposes of this paper. We recognize that in practice this write-only state will be enforced by a matter of policy, and that reads of this database are necessary for other functions (e.g., regulatory compliance with COPPA<sup>6</sup>).
- **Demographic and Behavioral Information**: Demographic and Behavioral information can be stored as part of the reporting and logging system and time to time inferencing can be performed over the data depending on how the system is set up. After each impression<sup>7</sup>, demographic and behavioral information can be updated for the GUID associated with the impression. If online learning is performed then requisite information from these databases is passed on to the modeled information knowledge base.
- **Inferencing Algorithm and Modeled Information**: As discussed in Section. 3.3, the inferencing algorithms can be of various types and have different information needs. Hence, modeled information can vary depending on the algorithm needs. Estimating the needs upfront always helps both the developers and the policy makers.

### 4.1 Example

As an example of the implementation of the “down-the-chain” notification policy, we talk about a system proposed by our research group at the University of Maryland. Our role in the OBA system is that of a *computer scientist* developing the fundamental algorithms. We specify our information needs at each phase of the algorithm which has made it easier for the actual implementors and designers to design an efficient system and to make available as many privacy options as possible for their system design.<sup>8</sup>

<sup>6</sup>Children’s Online Privacy Protection Act, 15 U.S.C. §§ 6501-06.

<sup>7</sup>A single view of a webpage is called an impression.

<sup>8</sup>This approach does not propose that roles earlier in the chain *require* various privacy features, only that they enable as many as possible and accurately document information needs so as to make the next role in the chain aware of the available options.

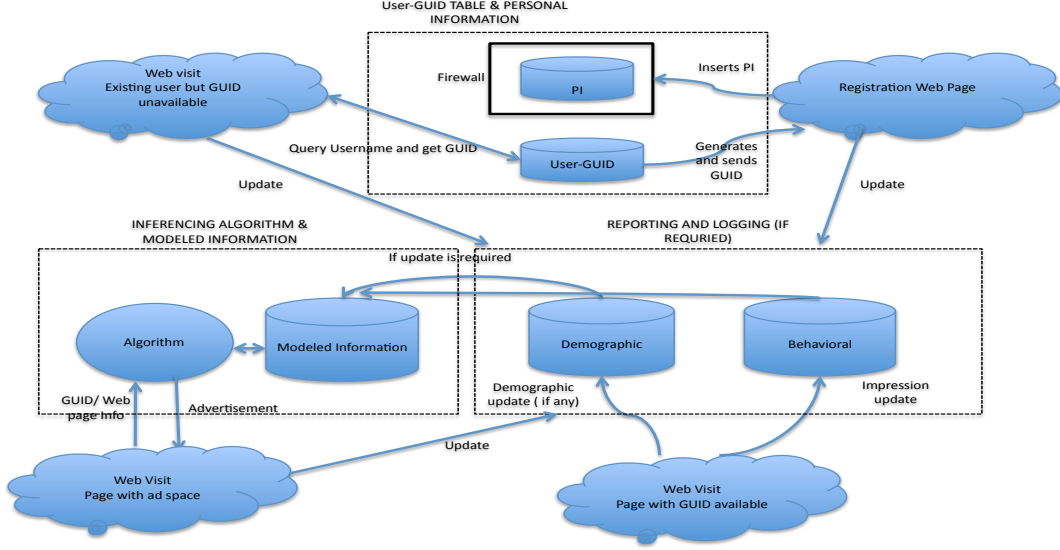


Figure 1: Sample Architecture for Online Advertising Server. PI refers to Personal Information and GUID is the Globally Unique Identifier.

$Ad_1$	#clicks	#impressions
$Ad_2$	#clicks	#impressions
...	...	...
$Ad_n$	#clicks	#impressions

Table 1: Modeled Information Phase 1

The online advertising system suggested by our group uses a form of “Contextual Bandit Algorithm” [10]. Contextual Bandit Algorithms perform online learning and are an advanced form of the multi-armed bandit problem [8]. Our algorithm is being developed incrementally and we have divided its development into 3 phases.

#### 4.1.1 Phase 1: Multi Armed Bandit Problem

Our system uses a bayesian inferecing based multi-armed bandit formulation to model the problem of online optimization in advertising. The multi-armed bandit problem can be formulated as follows: there is a bandit containing a set of arms  $(A_1, A_2, A_3, \dots, A_n)$ . Each arm has a success probability  $\theta_i$  associated with it which is unknown. A strategy needs to be decided to play the arms such that the total rewards obtained are maximized and the learning cost is minimized.

In the advertising domain, we can model each advertisement as an arm of a bandit and each impression of an advertisement as a play of an arm. The goal is to maximize the total number of clicks obtained by presenting more attractive advertising. Different multi-armed bandit strategies have been discussed in [11, 9]. These strategies are agnostic to the user and its information, hence the behavioral and static demographic data is not required. The only piece of information required for inferecing are the advertisement characteristics:

Segment <sub>1</sub>		
$Ad_1$	#clicks	#impressions
$Ad_2$	#clicks	#impressions
...	...	...
$Ad_n$	#clicks	#impressions

.....

Segment <sub>k</sub>		
$Ad_1$	#clicks	#impressions
$Ad_2$	#clicks	#impressions
...	...	...
$Ad_n$	#clicks	#impressions

Table 2: Modeled Information Phase 2

In the current form, there is no need for the system to store any information indexed or indexable to the user and only needs to store per advertisement information as shown in Table 1. The system will work fine if the necessary counters are incremented appropriately and the rest of the data is discarded (or not collected).

#### 4.1.2 Phase 2: Segmented Multi Armed Bandit Problem

It is a common understanding that different users behave differently and adding the user behavior can provide a boost in the advertising systems. Hence users can be segmented into groups and a separate model can be built for each group of the users. In our model, we use historical data consisting of demographic as well as behavioral data for segmenting users and then build a separate multi-armed bandit model for each segment as shown in Table 2.

When a new impression arrives with the GUID, a simple look-up is done to find out which segment the user belongs to and then the appropriate model is used. Otherwise, when a new user comes his information is updated in the behavioral



and demographic data and then his segment is decided using a pre-calculated formula for the segmentation.

Hence, in this phase the system needs to store only the information mentioned in Table 2 and the formula ( mapping) for user segmentation. All the other behavioral and demographic information can be discarded after the segmentation has been done.

#### 4.1.3 Phase 3: Contextual Bandit Problem

The preferences of the users change with time and dynamic systems can more accurately model individuals' preferences. So in its third phase, our system will also update the user level segmentation in an online fashion with each impression.

Much more information will need to be maintained and updated in this model since user context will be taken into account in an online manner in deciding which advertisement to display.<sup>9</sup>

Since *computer scientists* are one of the first in this chain of notification, we play an important role in fulfillment of the requirement of “down-the-chain” methodology.

## 5. OPEN QUESTIONS FOR DISCUSSION

This proposal represents preliminary discussions in the MIND Lab regarding the responsibilities we bear as computer scientists when developing new technologies. We have identified some questions that remain open issues in our proposal:

- Should developers also have to make public the data requirements (not just make that information available to purchasers/users of their systems)? Should these specifications be auditable?
- If there are different entities which supply different information pieces, then how should these information items be stored and managed? (e.g., if each of Yahoo and Google supply inputs to a behavioral marketer, should the marketer be required to publish the requirements of both systems?)
- Should advertisers (as different from operators of Ad Networks) have a role in the notification process?

## 6. CONCLUSION

We advocate that those who design and implement Online Behavioral Advertising systems should be required to document the information requirements of the algorithms and other technologies they build and maintain so that it is possible both to build systems that do not store more information than necessary and to enable others to audit whether more information is being stored than necessary. The main goal of this requirement is to enable as much consumer choice as possible.

Our “down-the-chain” notification policy will bridge the gap between the understanding of the functionality & requirements of the system amongst different role players and will make it easier for each player to work on his part. It will

<sup>9</sup>We are still working on the exact details of this system which will determine the sufficient information elements.

also provide a more coherent and useful view to the *end-consumer*.

## 7. REFERENCES

- [1] <http://advertising.yahoo.com/>.
- [2] <http://mediadecoder.blogs.nytimes.com/2009/08/05/an-interview-with-david-vladeck-of-the-ftc/>.
- [3] <http://mindlab.umd.edu/>.
- [4] <http://www.google.com/ads/>.
- [5] <http://www.privacybydesign.ca/>.
- [6] <http://www.teracent.com/>.
- [7] Consumers driving the digital uptake: The economic value of online advertising-based services for consumers. Interactive Advertising Bureau (IAB) Report, September 2010. [http://www.iab.net/insights\\_research/947883/](http://www.iab.net/insights_research/947883/).
- [8] P. Auer, N. Cesa-Bianchi, and P. Fischer. Finite time analysis of multi-armed bandit problem. *Machine Learning*, 27(2-3):235–256, 2002.
- [9] D. Chakrabarti, R. Kumar, F. Radlinski, and E. Upfal. Mortal multi-armed bandits. In *NIPS*, 2008.
- [10] L. Li, W. Chu, J. Langford, and R. E. Schapire. A contextual-bandit approach to personalized news article recommendation. In *WWW*, 2010.
- [11] S. Pandey and C. Olston. Handling advertisements of unknown quality in search advertising. In *NIPS*, 2006.
- [12] S. S. Shapiro. Privacy by design: Moving from art to practice. *Communications of the ACM*, 53(6):27–29, 2010.
- [13] J. Turow, J. King, C. J. Hoofnagle, A. Beakley, and M. Henessy. Americans reject tailored advertising and three activities that enable it, September 2009. SSRN: <http://ssrn.com/abstract=1478214>.
- [14] F. Williams. Internet privacy policies: A composite index for measuring compliance to the fair information principles, 2006. <http://www.ftc.gov/os/comments/behavioraladvertising>.

# DO NOT TRACK

## *An Attempt to Frame the Debate*

Hannes Tschofenig, Rob van Eijk

### I. INTRODUCTION

The Hypertext Transfer Protocol (HTTP), which was initially standardized with RFC 2068 [1], is a mostly stateless protocol. More sophisticated web applications that need to maintain state use the cookie concept, defined in RFC 2109 [2]. Cookies have found widespread usage in Web development and their current usage is being documented in [3].

Unfortunately, cookies have not only been used by web sites that the user explicitly wanted to connected to but instead it became common Web deployment practice to 'mash up' content from various other Web sites, including websites that provide advertising material. Over time the techniques for distributing information about users' web browsing behavior has become more sophisticated and researchers, such as the authors of [4], have described the state-of-the-art. The investigations indicate an increasing aggregation of user-related data.

The advertising industry was not inactive in light of the increasing concerns and have initiated various self-regulatory initiatives. [5] describes a few of these efforts and related attempts to block cookies.

With the publication of the preliminary Federal Trade Commission (FTC) privacy report [6] in December 2010, which followed a series of roundtable discussions, concerns about the development in the area of user tracking on the Web has gotten the attention of the industry. In discussions in early 2011, the FCC reiterated its support for the Do Not Track (DNT) concept and articulated several success criteria for DNT:

- 1) Implemented universally
- 2) Easy to use, find and understand
- 3) Persistent
- 4) Not only for use but also for collection
- 5) Effective and enforceable

In the meanwhile the European Commission has decided to tighten existing legislation by amending the e-Privacy Directive by the so-called 'EU Cookie Directive' [7]. Implementation of the directive into national

\*This position paper is a submission to the W3C Workshop on Web Tracking and User Privacy, 28/29 April 2011, Princeton, NJ, USA. Hannes Tschofenig is a senior standardization specialist at Nokia Siemens Networks. He is active at the Internet Engineering Task Force (IETF), co-chair of the Open Authentication Protocol (OAuth) working group, and a member of the Internet Architecture Board (IAB). Hannes was involved in the organization of the 'Internet Privacy Workshop (December 2010)' co-organized with MIT, W3C, IAB, and ISOC. He can be reached at Hannes.Tschofenig@nlnet.nl. Rob van Eijk is a Ph.D candidate at Leiden University, and employed at the Dutch Data Protection Authority. He can be contacted at r.vaneijk@blauw.com.

The content of the position paper represents the views of the authors and does not necessarily reflect the view of their employers, or any organization they authors are active in or are associated with.

law by European member states is required by May 2011. The directive requires end user consent to the storing of cookies on a computer.

Shortly after the publication of the preliminary FTC report industry players reacted by initiating standardization and implementation efforts. The IETF submission by Mozilla [8] suggested standardization of an HTTP header conveying a preference of the user not to be tracked (the "Do Not Track (DNT) header"). Microsoft submitted a similar contribution [9] to the W3C, which additionally contains a black list mechanism. In this document we focus on the Do Not Track header; the development of a black list is a largely orthogonal effort.

These two contributions and the Mozilla DNT contribution in particular raise a number of interesting challenges for the standardization community. In addition to the typical technical questions there are also questions about the interaction between the technical and the regulatory community.

In the sections below we list a couple of questions we find worthwhile to discuss.

## II. WHAT ARE WE TALKING ABOUT?

[8] attempts to define the scope of their work via the term 'tracking':

- **Tracking** includes collection, retention, and use of all data related to the request and response.

It seems to be natural to worry about the terminology and to scope the work appropriately.

### QUESTION #1: WHY CANNOT EXISTING TERMINOLOGY BE RE-USED?

Interestingly, Directive 95/46/EC (published October 1995) [10] defines terminology useful in this context. The relevant terms are:

- **Controller** shall mean the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific criteria for his nomination may be designated by national or Community law;
- **Processor** shall mean a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller;

Reusing this terminology also raises the question why the entire framework cannot be re-used altogether. This means that tracking will typically be treated in context of the interaction between the data controller and the data processor<sup>1</sup>.

[8] defines first party and third party in the following way:

- **First Party:** A first party is a functional entity with which the user reasonably expects to exchange data. In most cases the functional entity responsible for the web page a user has navigated to is the sole first party.
- **Third Party:** A third party is a functional entity with which the user does not reasonably expect to share data.

<sup>1</sup>To meet the page limit of this position paper we do not discuss the possibility to have multiple data controllers.

In the Directive 95/46/EC terminology [10] a first party most likely corresponds to the data controller and the third party to the data processor (even though an exact comparison between the two is not possible with the current definitions.

[8] describes a way to distinguish between the first party and the third party in an algorithmic way.

It is clearly a challenge to define such an algorithm to cover all cases. Given a definition companies will try to ensure that they fall under the first party category with the expectation that their responsibilities towards data subjects are reduced.

End users will not be able to understand the algorithmic definition. Data protection authorities have to work within the currently established legal framework to determine lawful processing. Any definition developed within a standards developing organization will not necessarily be accepted by the regulatory community.

**QUESTION #2: WHY IS AN ALGORITHMIC DESCRIPTION NECESSARY?**

### III. DID WE FORGET TO MENTION THE EXCEPTIONS?

The basic idea behind the list of exceptions is to point out that there are cases where the users preferences communicated via the Do Not Track indication are not honored. [8] attempts to define the following exceptions:

- 1) Tracking of users who have explicitly consented to tracking, such as by enabling a checkbox in a preferences menu on the first-party website of the tracking service.
- 2) Data obtained by a third party exclusively on behalf of and for the use of a first party.
- 3) Data that is, with high confidence, not linkable to a specific user or user agent. This exception includes statistical aggregates of protocol logs, such as pageview statistics, so long as the aggregator takes reasonable steps to ensure the data does not reveal information about individual users, user agents, devices, or log records. It also includes highly non-unique data stored in the user agent, such as cookies used for advertising frequency capping or sequencing. This exception does not include anonymized data, which recent work has shown to be often re-identifiable.
- 4) Protocol logs, not aggregated across first parties, and subject to a two week retention period.
- 5) Protocol logs used solely for advertising fraud detection, and subject to a one month retention period.
- 6) Protocol logs used solely for security purposes such as intrusion detection and forensics, and subject to a six month retention period.
- 7) Protocol logs used solely for financial fraud detection, and subject to a six month retention period.

The preliminary FTC privacy report [6] also touched this topic with an attempt to simplify privacy notices to data subjects by first parties. The FTC staff solicited comments on what is considered "commonly accepted practice" for which companies should not be required to seek consent once the consumer elects to use the product or service in question. The report itself lists the following items:

- **Product and service fulfillment:** Websites collect consumers contact information so that they can ship requested products. They also collect credit card information for payment. Online tax calculators and financial analysis applications collect financial information to run their analyses for customers.
- **Internal operations:** Hotels and restaurants collect customer satisfaction surveys to improve their customer service. Websites collect information about visits and click-through rates to improve site navigation.

- **Fraud prevention:** Offline retailers check drivers licenses when consumers pay by check to monitor against fraud. Online businesses also employ fraud detection services to prevent fraudulent transactions. In addition, online businesses may scan ordinary web server logs to detect fraud, deleting the logs when they are no longer necessary for this purpose. Stores use undercover employees and video cameras to monitor against theft.
- **Legal compliance and public purpose:** Search engines, mobile applications, and pawn shops share their customer data with law enforcement agencies in response to subpoenas. A business reports a consumers delinquent account to a credit bureau.
- **First-party marketing:** Online retailers recommend products and services based upon consumers prior purchases on the website. Offline retailers do the same and may, for example, offer frequent purchasers of diapers a coupon for baby formula at the cash register.

**QUESTION #3:** ARE SPECIFICATIONS FROM STANDARDS DEVELOPING ORGANIZATIONS THE RIGHT PLACE TO DEFINE THIS TYPE OF POLICY?

**QUESTION #4:** HOW LIKELY IS IT THAT SUCH A POLICIES WILL VARY BETWEEN JURISDICTIONS?

#### IV. HOW DOES THE ENFORCEMENT WORK?

The concept of the DNT indication inherently relies on the idea that bad actors, who do not adhere the user's DNT preferences, get prosecuted via the legal framework. There are no technical enforcement mechanisms built-in. There is problem by itself with such an approach.

First, there is the question of how users (or more realistically researchers, etc. on behalf of users) detect failure to comply. Data sharing can always happen in the background without exposing any traces to end devices. A second aspect is whether the conveyed preference in a header is enough basis for enforcement actions, particularly if the preference had been sent over an insecure channel that allows intermediaries (such as proxies) to modify settings.

**QUESTION #5:** HOW DO WE ENVISION MISBEHAVIOR TO BE DETECTED?

**QUESTION #6:** DOES A SET HEADER PROVIDE ENOUGH BASIS FOR ENFORCEMENT BY DATA PROTECTION AUTHORITIES?

#### V. CONCLUSIONS

In this position paper the authors raise a number of questions relevant to the ongoing standardization debate. From the perspective of the authors existing terminology shall be re-used, an algorithmic definition of first party vs. third party is not needed, exceptions must not be defined by standards developing organizations but rather left to the regulatory community and will vary among jurisdictions.

We therefore suggest to focus the standardization work on developing technical building blocks that support the existing and evolving regulatory framework. A discussion about the layer in the protocol stack (as well as the appropriate header field) at which the preference indication should be conveyed is within the realm of standards organizations to decide. The needed implementation complexity has to be taken into consideration. The responsibilities for desired behavior have to be clearly articulated. Another technical question that may

need discussion is whether this DNT capability should only be restricted to HTTP but be re-applied to other protocols, such as email, SIP, or XMPP.

#### REFERENCES

- [1] R. Fielding, J. Gettys, J. Mogul, H. Frystyk, and T. Berners-Lee, "Hypertext Transfer Protocol – HTTP/1.1," Jan. 1997, RFC 2068, Request For Comments.
- [2] D. Kristol and L. Montulli, "HTTP State Management Mechanism," Feb. 1997, RFC 2109, Request For Comments.
- [3] A. Barth, "HTTP State Management Mechanism," Mar. 2011, IETF draft (work in progress), draft-ietf-httpstate-cookie-23.txt.
- [4] B. Krishnamurthy and C. Wills, "Privacy diffusion on the web: a longitudinal perspective," in Proceedings of the 18th international conference on World wide web, ser. WWW '09. New York, NY, USA: ACM, 2009, pp. 541–550. [Online]. Available: <http://doi.acm.org/10.1145/1526709.1526782>
- [5] A. Cooper and H. Tschofenig, "Overview of Universal Opt-Out Mechanisms for Web Tracking," Mar. 2011, IETF draft (work in progress), draft-draft-cooper-web-tracking-opt-outs-00.txt.
- [6] Federal Trade Commission, "Protecting Consumer Privacy in an Era of Rapid Change; A Proposed Framework for Businesses and Policymakers," Dec. 2010, Report available for download at <http://www.ftc.gov/opa/2010/12/privacyreport.shtm> (Apr. 2011).
- [7] European Commission, "Directive 2009/136/EC of the European Parliament and of the Council of 25 November 2009 amending Directive 2002/22/EC on universal service and users rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws," Nov. 2009, Official Journal L 337/11, 18/12/2009.
- [8] J. Mayer, A. Narayanan, and S. Stamm, "Do Not Track: A Universal Third-Party Web Tracking Opt Out," Mar. 2011, IETF draft (work in progress), draft-mayer-do-not-track-00.txt.
- [9] A. Zeigler and A. Bateman and E. Graff, "Web Tracking Protection," Feb. 2011, Contribution available at <http://www.w3.org/Submission/web-tracking-protection/> (Apr. 2011).
- [10] European Commission, "Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data," Oct. 1995, Official Journal L 281, 23/11/1995 P. 0031 - 0050.

**From:** "Tanya Tan" <tan@valueclick.com>  
**Subject:** ValueClick, Inc.'s Position Paper to Participate in WC3 Workshop on Web Tracking and User Privacy 4/28-4/29  
**Date:** 1 April 2011 03:08:39 GMT+02:00  
**To:** <team-privacyws-submit@w3.org>  
**Cc:** "Scott Barlow" <sbarlow@valueclick.com>, "Peter Wolfert" <pwolfert@valueclick.com>  
**Archived-At:** <<http://www.w3.org/mid/CDA2F765A586124DAD93C09E110B79A902266F85@LA-EXCLUST01.corp.valueclick.com>>

---

To the World Wide Web Consortium (W3C) -

ValueClick, Inc. is one of the world's largest and most comprehensive online marketing service companies and was most recently named by Paid Content as #31 of the 50 most successful digital companies in the United States (see <http://paidcontent.org/list/page/the-most-successful-digital-companies/P34/>).

We have been a participant in the online advertising space for thirteen years and have four primary business segments - Affiliate Marketing, Media, Owned & Operated Websites and Technology. In the affiliate marketing space, our Commission Junction is a global leader in the online advertising channel for affiliate marketing. In the media space, ValueClick Media was founded in 1998 and was among the first online advertising networks and an early pioneer of performance-based pricing models. In the owned and operated website space, we have multiple website properties including [www.smarter.com](http://www.smarter.com), [www.couponmountain.com](http://www.couponmountain.com) and our most recent acquisition, [www.investopedia.com](http://www.investopedia.com). In the technology space, we have Mediaplex, Inc. that offers cross-channel advertising technology and service solutions, including ad serving services.

We understand the importance of data privacy and have committed significant resources towards embedding privacy in our organization through our active participation in self-regulatory organizations such as the Network Advertising Initiative (NAI), the Interactive Advertising Bureau (IAB) and Internet Advertising Sales Houses (IASH). As such, our large and diverse role in the online advertising marketplace and our commitment to user privacy make us a material stakeholder in the discussion regarding web tracking and user privacy (April 28-29) and we would like to be a participant in such workshop.

Please let us know if you have any questions. Thank you.

Tanya M. Tan  
Assistant General Counsel and VP, Privacy & Legislative Affairs  
ValueClick, Inc.  
30699 Russell Ranch Road, Suite 250, Westlake Village, CA 91362  
phone: 818.575.4739 | fax: 818.575.4904  
email: [ttan@valueclick.com](mailto:ttan@valueclick.com)

This email and any files included with it may contain privileged, proprietary and/or confidential information that is for the sole use of the intended recipient(s). Any disclosure, copying, distribution, posting, or use of the information contained in or attached to this email is prohibited unless permitted by the sender. If you have received this email in error, please immediately notify the sender via return email, telephone, or fax and destroy this original transmission and its included files without reading or saving it in any manner. Thank you.

## **Letter of Interest - W3C Workshop on Web Tracking and User Privacy**

Ian Plunkett - Director of Eng. VizScore Inc.

April 26, 2011

### **Introduction**

VizScore is an early stage startup in the email marketing space. Email marketing is most effective to the extent that data about individuals is collected as they leave information about themselves across the web - usually email addresses and often times more intimately identifiable details. Using these data to support marketing initiatives presents a number of unique challenges. Capturing information about browsing habits can be beneficial to consumers and marketers alike as it can enable increasingly relevant marketing communication. Marketers can present a more personalized experience that can lead to an increased ROI and consumers can be directed to products and services they are more likely to want. On the flipside, most consumers do not understand how and where data they have implicitly and explicitly provided are distributed across the web. This lack of understanding often leads to a general fear of data collection. A combination of education and technical standards will become increasingly important in maintaining a healthy balance that protects the identities of consumers on the web and allows for the best possible experience online.

### **Rising Influence of Real Identities across the Web**

The rising notion that a user's online persona should be directly tied to their real life identity will complicate efforts to assuage concerns over tracking users across the web. To take an example from the fairly ubiquitous Facebook, we have "Like" buttons, Facebook Connect and Facebook Comments implemented across an ever-growing number of sites (Google and others have similar services.) Analyzing data produced by these widgets creates a fairly comprehensive user behavior profile. As these types of services gain traction, the user experience across the web becomes more and more reliant on personalized services. If the user chooses to opt out of such services, they will experience a degraded version of the web. Using HTTP headers to add "Do Not Track" features to websites will result in two-tiered systems. Many website developers may choose to entirely block visitors who enable such features because the development cost and subsequent ROI will not justify the effort. This makes initiatives to introduce code implementing "Do Not Track" features less attractive and will most likely hamper widespread adoption. Similarly, users that are not tech savvy may have difficulty even understanding the ramifications of allowing themselves to be tracked online and may prefer that things "just work."



## **Online Data Retention and Theft**

Data theft presents a persistent problem for any repository of personally identifiable information. Recently, crackers were able to breach Epsilon, a leading email marketing company, and steal personally identifiable information on an unknown (presumably large) number of people. The attack against Epsilon is not an isolated incident and any organization that tracks users online should have comprehensive security policies to minimize the risk of data breaches. Companies and organizations collecting behavioral information should also have clearly outlined and publicly available policies that describe what type of information they retain on individuals. Further, they should provide a path for consumers to request that their data be removed from said systems.

## **Conclusion**

Through a combination of consumer/user education and clear and open policies, we can align marketing interests and consumer privacy protection concerns. Behavioral tracking on the web can be beneficial to both consumers and marketers. The industry should strive to keep the public informed of how they protect personal data and how they use those data to enhance their marketing communications and user experience.

# Toward An Empirical Investigation of Usability and Effectiveness of Do-Not-Track Tools

Position Paper Submitted to  
W3C Workshop on Web Tracking and User Privacy

March 2011

Yang Wang, Lorrie Faith Cranor  
wang@cs.cmu.edu, [lorrie@cs.cmu.edu](mailto:lorrie@cs.cmu.edu)  
CyLab Usable Privacy and Security Lab (CUPS)  
Carnegie Mellon University

## Introduction

Internet users have repeatedly expressed strong aversion to behavioral advertising and online tracking, raising concerns about privacy [1][2]. Efforts to address these privacy concerns focus on giving users privacy controls. However, for controls to be effective they must be usable and work as advertised.

A key element of industry self-regulations is to allow users to opt out of behavioral advertising. A number of opt-out mechanisms have been designed and deployed including but not limited to Network Advertising Initiative (NAI) opt-out tool, Google advertising cookie opt-out tool, Targeted Advertising Cookie Opt-out (TACO) tool, and opt-out tools in web browsers such as IE9 and Mozilla Firefox. However, we have not seen any study that investigates the effectiveness of this opt-out approach.

We set out to answer this important, timely, and practical question – *do these opt-out mechanisms actually work for ordinary users?* We plan to conduct a series of studies to empirically investigate their effectiveness in protecting users from targeted ads and online tracking and identify the causes of any deficiencies. This position paper describes our research plans.

## Research Focus

Our research will focus on the following four areas:

- Taxonomy of opt-outs
- Usability of opt-out tools
- Effect of opt-out tools on behavioral advertising
- Effect of opt-out tools on online tracking

There are a wide variety of websites that claim to support these opt-out options. For instance, NAI currently has 66 members that support the NAI opt-out tool. However, it is not clear whether these sites all interpret the opt-out the same way. We have begun to classify these sites, both in how they describe their response to the opt-out, and in how they actually respond to the opt-out, e.g., do they place further cookies on the user's browsers, and if so, what information do the cookies contain?

Because the opt-out process involves users' actions, the second area of this research focuses on the usability of these tools. If the tools are not usable, users will not be able to gain benefits from using the tools regardless of their intended functionalities.

Opt-out tools will be ineffective if ad networks choose not to respect the opt-out set by the tools, or if the tools themselves do not behave exactly as designed (e.g., they do not actually set an opt-out cookie or that cookie gets deleted). We will treat the behavioral advertising system as a black box and devise a systematic experiment scheme to observe and deduce the effect of these opt-out tools on behavioral advertising.

It is important to note that even if sites do not provide targeted ads to users, they could still track them online. Therefore, we also plan to examine how these tools might affect sites' online tracking practices.

## **Research Methodology**

We will select a set of representative opt-out tools. This set will cover a variety of ways in which these tools have been designed and implemented such as browser plug-ins and native browser features (both on computers and mobile devices), DNT headers, and websites.

### **Usability of opt-out tools**

To evaluate the usability of these opt-out tools, we will conduct a heuristic evaluation (a form of expert review) with a few usable privacy experts and a lab usability study with ordinary Internet users. Then each participant will be randomly assigned to use one of the selected opt-out tools, and asked to install the tool on a lab machine. We will give them the same set of tasks such as opting out from tracking by a certain site. We will ask them to think aloud while they install the tool and perform the tasks. At the end of study, we will ask them to fill out a subjective satisfaction survey and briefly interview them about their experience with the tool. We will video tape (without recording their face) the study and screen record their interactions with the tool. In addition, we will measure (1) how long does it take each participant to complete a task (e.g., install the tool)? (2) task success rate, and (3) their subjective assessment of the tool on a Likert scale.

For qualitative data such as interviews, we will transcribe them and identify potential usability problems. For quantitative measurements such as task performance time and task success rate, we plan to analyze the data using statistical tools.

### **Effectiveness of opt-out tools**

Our scheme to test the effect of an opt-out tool is inspired by Guha et al. [3] and our work examining Flash LSO re-spawning behavior [4]. We set up two web browser instances with only one difference (one instance enables the tool while the other does not) on the same machine.

We script both browser instances to visit the same set of “learning” websites and a destination site (e.g., a news site). We would choose “learning” websites that have clear themes, e.g., sites of baseball teams, so that user interests can be learned. For instance, one can reasonably assume that if a user visits baseball team websites frequently, this user is likely to be interested in baseball and a targeted ad system may display ads about baseball. The scripts would also keep track of all the cookies and ads each browser receives along the way. We will check how the received ads match up with the themes or characteristics of the visited sites to assess whether they are targeted ads. We then compare the two sets of ads to see how similar they are. After controlling for random noise in ad selection, generally speaking, the more difference between the two sets of ads, the more effect the opt-out tool has.

## Conclusions

The underlying model of these behavioral advertising opt-out tools relies on an important assumption that users can easily understand and use these tools to express their opt-out preferences. We have seen too many cases where brilliant security and privacy technologies are simply not usable and thus unused (e.g., PGP [5]). Therefore, usability is a key factor to the success of such tools. To our knowledge, no systematic usability evaluation has been conducted on these opt-out tools. Our study is likely to be the first. It has important research and practical value. We expect to identify major usability problems from this study and to create guidelines to help design better usability in such tools.

Our simulation experiment on the effect of these tools is based on actual system responses as if a user is browsing the Web. It is considerably more objective and reliable than people’s self-reported attitudes and behavior. The results, either positive or negative, will have substantial implications to the industry, regulators and privacy technologists.

## Reference

- [1] J. Turow, J. King, C. J. Hoofnagle, A. Bleakley, and M. Hennessy, “Americans Reject Tailored Advertising and Three Activities That Enable It,” *SSRN eLibrary*, 29-Sep-2009. [Online]. Available: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1478214](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1478214). [Accessed: 08-Jan-2010].
- [2] Aleecia M. McDonald and Lorrie Faith Cranor, “Beliefs and Behaviors: Internet Users’ Understanding of Behavioral Advertising,” presented at the The 38th Research Conference on Communication, Information and Internet Policy, 2010.
- [3] S. Guha, B. Cheng, and P. Francis, “Challenges in Measuring Online Advertising Systems,” in *Proceedings of the 10th ACM SIGCOMM conference on Internet measurement conference*, 2010.
- [4] Aleecia M. McDonald and Lorrie Faith Cranor, *A Survey of the Use of Adobe Flash Local Shared Objects to Respawn HTTP Cookies*. 2011.
- [5] A. Whitten and D. Tygar, “Why Johnny Can’t Encrypt: A Usability Evaluation of PGP 5.0.,” in *Ninth USENIX Security Symposium*, 1999.

# Identifying and Preventing Conditions for Web Privacy Leakage

Craig E. Wills

Computer Science Department  
Worcester Polytechnic Institute  
Worcester, MA 01609

## 1 Position Statement

Beyond tracking and proposals to limit it, previous work has observed the *leakage* of private user information to a variety of *third-party aggregators* on the Web via a range of first-party Web sites [3, 5]. These first-party sites include both traditional and mobile Online Social Networks (OSNs) as well as non-OSNs where users register and supply personal information as part of setting up an account.

My position is that we not only need to be concerned with tracking, but also need to identify the conditions under which leakage occurs and work to prevent one or more of these conditions. My work puts me in a position to identify these conditions and point at steps that can be taken by sites and users to prevent them. I focus my attention on leakage to third parties that are present on first-party sites because unlike first-party sites they have the means to observe and *link* the behavior of, and information about, users across multiple first-party sites. I show that while users can take some actions, first-party sites are in the best position to prevent this leakage of private information about their users.

## 2 Leakage Conditions

Identifying the set of necessary conditions for a problem to occur has been done for other domains. One classic example is for deadlock where [1] identified four conditions that must be present for deadlock to occur and observed that deadlock can be prevented by negating one or more of these necessary conditions. I take a similar approach for the problem of Web privacy leakage where I identify necessary conditions for privacy leakage, examine specific circumstances where these conditions prevail to cause leakage and look at techniques for prevention of leakage by negating one or more of these necessary conditions.

Based upon my own work, I identify three necessary conditions under which I observe the leakage of private information to a third-party aggregator:

1. A user makes information about themselves available to a first-party site. This could be private information such as name or email address, information about their zip code, such as provided through a “store finder” service on a shopping site, or a more-precise latitude/longitude location, such as through a mobile device location.
2. The first-party site receiving this information exposes it in a manner that is visible via HTTP transactions. This exposure is typically through a HTTP request header.
3. A third-party aggregator is present on the first-party site and some amount of user-provided information is made available to the third party as part of a HTTP transaction with the third-party server.

### 3 Instances of Leakage Conditions

Based on past and current work, I have observed five instances where these three conditions for leakage are realized. While these five instances are not necessarily exhaustive, they do represent the range of leakages that I have observed across both OSN and non-OSN Web sites. In each of the following cases I assume that a user has already provided potentially private information (Condition 1) to a first-party site and show both how the first party exposes this information (Condition 2) in a HTTP transaction to a third-party aggregator (Condition 3). I show representative examples in each case, but intentionally use generic first- and third-party server names to focus on the nature of leakage rather than the specific parties. I have observed a number of instances of all types shown.

#### 3.1 Transmission of User Input via Request-URI

Users provide information about themselves to a first-party site when they edit their user profile or enter terms as part of a search. If this information is transmitted to the first party via the Request-URI then it may be leaked to a third-party in one of two related ways. First it may be leaked to a third-party server via the HTTP `Referer` header if a third-party object is present on the page with the information in the Request-URI. This situation is shown in the following where a zip code is included in the Request-URI by the first party and subsequently leaked by the `Referer` header in a request to `tracker.thirdparty.com`.

```
GET http://tracker.thirdparty.com/params...
Referer: http://www.firstparty.com/...zip=12201...
```

A variant of this leakage occurs when the *next* first-party page a user visits contains third-party JavaScript code that retrieves the referring URL via the JavaScript API and subsequently passes this URL (containing the private information) to the third-party server.

```
GET http://track.thirdparty.com/...
  referer=http://www.firstparty.com/...zip=12201...
Referer: http://www.firstparty.com/nextpage...
```

#### 3.2 Inclusion of Private Information in Page Title

Another example of leakage occurs when first-party sites expose private user information in the title of a Web page, which is then obtained by a third-party script via the JavaScript API. A common example of this type of leakage is when a user's name is put in the title of the user's profile page on a site. This name is subsequently leaked when third-party JavaScript code executes, obtains the page title contents as part of execution and returns it to the third party as part of the Request-URI. Note the example also shows the user's identifier for the site being leaked in the `Referer` header.

```
GET http://tracker.thirdparty.com/...title=John Doe profile...
Referer: http://www.firstparty.com/profile/123456789...
```

#### 3.3 Leakage via First-Party Cookies to Hidden Third-Party

Some sites store private information about the user, such as name or email address, in site-specific first-party cookies. Leakage of this private information occurs when these sites also employ what is referred to as *hidden third-party* servers where a given server looks like it belongs to a first-party domain, but actually belongs to a third party [4]. An example of this type of leakage is illustrated below where email and full

name are passed to `thirdparty.firstparty.com` because the cookies containing these values are associated with the `firstparty.com` domain and the browser interprets this third-party server as being part of the first-party domain.

```
GET http://thirdparty.firstparty.com/...
Referer: http://www.firstparty.com/...
Cookie: ...e=jdoe@email.com&f=John&l=Doe...
```

### 3.4 First-Party Information Used to Populate Third-Party Request-URI

This leakage occurs when information available to the first party is used to populate parameters of a third-party Request-URI. The following example shows such leakage where a user's age, gender and zip code are leaked directly to `tracker.thirdparty.com`. This example demonstrates explicit leakage of first-party information to the third party.

```
GET http://tracker.thirdparty.com/...age=30&gender=M&zip=12201...
Referer: http://www.firstparty.com/...
```

### 3.5 Information POSTed to Third Party

The final type of leakage was observed in [5] where it was noted that smart phone applications are able to obtain information about a user's device and transmit this information to a third party. The following example shows that a third party is passed the device identifier and latitude/longitude via the API available to the first-party application.

```
POST http://tracker.thirdparty.com/
User-Agent: firstpartyapp/2.2.0 CFNetwork/459

id=IPHONE-UDID,lat=20.00,lon=-70.00
```

## 4 Leakage Prevention

As noted in [3], third parties receiving private information could filter what is received and not use it. However I believe the right approach is to ensure that third parties do not even receive the information so there is no question on whether or not they are in a position to use it. That leaves two entities—the user and the first-party site—to prevent the leakage of private information by negating one of the three conditions for leakage as defined in Section 2. To illustrate I describe possible actions available to each entity and how each action specifically negates one of the three conditions as well as which leakage instances in Section 3 are prevented.

### 4.1 User Actions

The simplest approach available to a user is to negate Condition 1 by not providing any private information to a first-party site—a site cannot leak what it does not know. However, creation of an account on a site may be a prerequisite for using the site, such as for an OSN, or the creation of an account may be needed to access valued functionality. In examining a variety of sites on what information is *minimally required* for registration, I found that 95% require an email address while roughly half require some combination of full name, date of birth, zip code and gender.

Given that users do not control first-party site exposure of information, further user-controlled prevention must be done by negating Condition 3. I identify three such user-based actions.

1. One approach that limits leakage via the `Referer` header, as shown in Section 3.1, is to control its use via browser settings. However, Internet Explorer and Safari do provide a setting to control when the `Referer` header is sent and while Firefox and Chrome browsers do provide such a setting, it is disabled by default and requires technical knowledge to enable it [7].
2. Another action available to users for prevention of leakage is to disable JavaScript execution through browser settings or do so selectively via a tool such as NoScript [6]. This action eliminates leakage shown in Sections 3.1 (second example), 3.2 and 3.4 (when information population is done via JavaScript variables). Unfortunately disabling JavaScript execution can negatively affect page quality and cause pages to break [2].
3. Users can use an ad blocker to block all requests to known third-party aggregators. This action is effective when the set of third parties can be identified and negates Condition 3 for examples in Sections 3.1, 3.2 and 3.4. A survey of users on actions taken for privacy protection found that 56% of respondents reported having used ad blockers [8]. However this approach requires that the set of known third parties be maintained and blockage of all hidden third-party servers, as shown in Section 3.3, is difficult.

Other actions available to users regarding blocking cookies or opting out from third-party cookies may inhibit tracking and linking of user records, but do not negate any of the conditions for leakage of information.

## 4.2 First-Party Actions

Unlike users, first-party sites control what information is exposed in HTTP transactions and can therefore prevent inadvertent leakage by negating Condition 2 via a number of actions.

1. Leakage of the type shown in Section 3.1 can be prevented by not passing the user input via the Request-URI, but by using a HTTP POST method and passing the input as part of the body of the request. With this approach the private information is not exposed in the Request-URI and any third parties will not obtain the information via the `Referer` header.
2. As noted in [3], Facebook uses a variant of this approach by putting a user's identifier after a '#' symbol in the Request-URI. Information after this symbol is not included by browsers in generating the `Referer` header.
3. First-party sites can prevent the "page title" leakage described in Section 3.2 by not putting private information in a Web page title, but rather put it in the contents of the page itself. This approach prevents access to the private information via the JavaScript API.
4. First-party sites can also prevent leakage to hidden third-party servers (Section 3.3) either by not using such servers or alternately changing how cookies are set for a first-party domain. Rather than associate cookies with the domain `firstparty.com`, they should be associated with `www.firstparty.com` so that hidden third parties within the domain (e.g. `thirdparty.firstparty.com`) do not have access to the cookies and therefore cannot obtain their contents.
5. An alternate approach for preventing leakage of the types shown in Sections 3.1 and 3.3 is for first-party sites to hash the private information so that its value is not readable by a third party that may receive the information.



These first-party actions prevent inadvertent leakage of private information by first-party sites, but these actions do not prevent the leakage described in Sections 3.4 and 3.5. The leakage in Section 3.4 shows cooperation by the first-party site to populate the Request-URI so leakage prevention requires the first party to cease such cooperation. The leakage in Section 3.5 is not directly in control of the first-party site and can only be prevented by the first-party application no longer making use of the given third party.

## 5 Summary

In this position statement I have identified three necessary conditions for private user information made available to a first-party Web site to be leaked to a third-party aggregator. I go on to provide five specific instances of where leakage occurs and show how this leakage can be prevented through a number of actions available to users as well as first-party sites. I believe an understanding of how leakage of private information occurs on the Web is a necessary step in developing technology to help prevent it.

## References

- [1] E.G. Coffman, Jr., M.J. Elphick, and A. Shoshani. System deadlocks. *ACM Computing Surveys*, 3(2), June 1971.
- [2] Balachander Krishnamurthy, Delfina Malandrino, and Craig E. Wills. Measuring privacy loss and the impact of privacy protection in web browsing. In *Proceedings of the Symposium on Usable Privacy and Security*, pages 52–63, Pittsburgh, PA USA, July 2007.
- [3] Balachander Krishnamurthy and Craig E. Wills. On the leakage of personally identifiable information via online social networks. In *Proceedings of the Workshop on Online Social Networks in conjunction with ACM SIGCOMM Conference*, August 2009.
- [4] Balachander Krishnamurthy and Craig E. Wills. Privacy diffusion on the web: A longitudinal perspective. In *Proceedings of the World Wide Web Conference*, pages 541–550, Madrid, Spain, April 2009. ACM.
- [5] Balachander Krishnamurthy and Craig E. Wills. Privacy leakage in mobile online social networks. In *Proceedings of the Workshop on Online Social Networks*, June 2010.
- [6] Noscript. <http://noscript.net/>.
- [7] Christopher Soghoian. Slight paranoia: It is time for the web browser vendors to embrace privacy by default, October 17 2010. <http://paranoia.dubfire.net/2010/10/it-is-time-for-web-browser-vendors-to.html>.
- [8] Craig E. Wills and Mihajlo Zeljkovic. A personalized approach to web privacy—awareness, attitudes and actions. *Information Management and Computer Security*, 19(1), 2011.

## W3C Proposal – DAA DNT Hybrid

### Do Not Track Headers and CLEAR Ad Notice

Most major web browser vendors recently released features aligned with emerging regulatory calls for a “Do Not Track” solution to online behavioral advertising. Major web browsers recently released features that align with calls for a “Do Not Track” solution to online behavioral advertising – although each company has taken a different approach to tackle the challenge. A better outcome for consumers is to converge on a single approach to exercising DNT choices to online behavioral advertising through web browser controls to reduce confusion and to better align the user experience with the consistency of the CLEAR Ad Notice program managed by the Digital Advertising Alliance (DAA). In short, it is proposed that web browser vendors align behind a single Do Not Track approach to increase consumer awareness through education and exposure to these features (additive vs. distractive).

It’s important to keep in perspective that advertising fuels the vast majority of free content and experiences available to consumers across the Internet today. The sites who invest the time, energy, employees, and technology to provide these free experiences must be equal partners in the conversation about these standards. All stakeholders should seek to find a balance to consumer privacy protections and a publisher’s ability to monetize their efforts.

#### DAA/DNT Hybrid Solution

As evidenced by the pains of removing IE6 from general use, it will take users time to upgrade to versions of web browsers that support a consistent, cohesive DNT solution. As currently implemented, DNT headers do not provide granular consumer control over their experience and their ability to express greater options for the brands they trust.

CLEAR Ad Notice was conceived and deployed to provide consumers with more granular information and choice in direct association with the ad they are seeing at that moment. In combination with CLEAR Ad Notice many participants in the advertising ecosystem are also launching detailed transparency and control tools to manage their advertising interests.

With that in mind, a hybrid DAA / DNT Header approach should be adopted to embrace simplified, persistent user controls native to the web browser and merge these with the mature opt-out programs already available to consumers which in turn are mated with maturing transparency mechanisms available through the Advertising Options Icon (CLEAR Ad Notice).

- **DAA:** Provides **transparency** and **granular choice** to users through existing solutions (backwards compatible)
- **DNT Header:** Provides the ability for opt-outs to be **persisted** and **evaluated/enforced**

#### How would this work?

Submitter: Shane Wiley, Sr. Director – Privacy & Data Governance, Yahoo!, wileys@yahoo-inc.com

- Setting Choice: User can set choice either through browser UI (DNT) or through Opt-Out pages (individual or group pages like NAI and AboutAds) available through CLEAR Ad Notice
  - Opt-Out signals are honored whether from the DNT Header or from the Opt-Out Cookie
  - DNT signal with a different value is sent to domains that are “trusted” (see “DNT Exceptions” below)
- Response to Choice: Once a DNT signal is received, the domain responds with a header response for the domain so the browser, the user, and interested 3<sup>rd</sup> parties can confirm the signal was received and appropriately accepted
  - One of two values should be returned:
    - acknowledged but not honored (see “DNT Exceptions”);
    - or, honored.
  - A DNT cookie should be set to allow for external auditing of consumer choice (the DNT signal itself will remain persisted within the browser UI – the cookie is merely for transparency and audit purposes)
    - Modify existing opt-out cookies with a new DNT value;
    - Or, develop an industry DNT cookie for all parties to set to simplify external auditing

### **DNT Exceptions**

To provide consumers with a level of choice (versus an “all or nothing” proposition), it will be important for users to be able to express exceptions to a DNT request. This approach also allows for the “quid pro quo” relationship between publishers and consumers to be expressed in a transparent and editable manner (allowing a user to change their mind at any time).

- Exceptions could be single entries or lists (users should have the option to view the entire list prior to agreement for its application)
- Entries should be expressed as a simple core domain name to simplify the experience for users (for example – publisher123.com, adnetwork345.com, or contentprovider567.com).
- While not necessary it would be beneficial if DNT Exception Lists could be subscription based (“off” by default) to reduce the nuisance to consumers as publishers engage in new 3<sup>rd</sup> party relationships.
- 1<sup>st</sup> parties should receive a signal if one or more of the 3<sup>rd</sup> parties available on their property have been blocked. This will provide the publisher with the option to provide a different (possibly reduced) experience to the user or for the user to provide an exception to gain access to free content.

### **Definition of “Do Not Track”:**

The W3C should not attempt to define what DNT means and instead leave this definition to be created by policy development and self-regulatory groups in partnership with consumer advocates and regulators.

This proposal has been developed to be implementable regardless of the DNT definition. That said, this submitter believes it would be most appropriate for industry to continue to maintain current, consistent

industry definitions. As such, at a high-level the scope of the **Do Not Track signal should be equal to the handling of today's behavioral advertising opt-out.**

Notably:

- **Do Not Profile:** The browser activity should not be added to a “profile” of the cookie – this extends to site retargeting efforts which cross non-commonly branded sites
- **Do Not Target:** The browser/device should not be targeted with online behavioral advertising (OBA)
- **Operational Needs:** Standard data collection for operational needs such as impression counting, frequency capping, and fraud detection/defense efforts is still supported.
- **1<sup>st</sup> Party:** Data collection and personalization activities provided by a 1<sup>st</sup> party are not subject to DNT. This extends to 3<sup>rd</sup> parties providing services only to the 1<sup>st</sup> party domain on their behalf and not developing cross non-commonly branded site OBA profiles.
- **Analytics:** Anonymous data necessary for basic reporting of impressions, clicks, and conversions should be maintained (not used to alter future browser experiences - outside of fraud defense)

### **Honoring User Preferences**

As multiple systems may be setting, sending, and receiving DNT and/or Opt-Out signals at the same time, it is important to ensure publishers, advertisers, ad networks, and web browser vendors consistently honor user choices in circumstances where “mixed signals” may be received.

- **No DNT Signal / No Opt-Out:** Browser / device is not opted-out
- **DNT Signal / No Opt-Out:** Browser/device is opted-out
- **Opt-Out / No DNT Signal:** Browser/device is opted-out
- **Opt-Out / DNT Exception:** Exception is honored (browser/device is not opted-out)

### **Conclusion**

Yahoo! strongly supports the standards development process and is submitting these recommendations in the hope that vigorous, enlightened, respectful debate ensues to drive consensus towards a solution that meets the needs of consumers, publishers, advertisers, and the parties that support each.

# Adobe Position Paper on Privacy and Tracking

---

**March 24, 2011**

**Submitted to the W3C in Anticipation of Participating in the  
W3C Workshop on Web Tracking and User Privacy  
By MeMe Jacobs Rasmussen, VP, Chief Privacy Officer**

## Introduction

Adobe believes that the W3C workshop on web privacy and tracking represents an important first step in an examination of a very complex and growing issue that affects all of the participants of the World Wide Web. Rarely has there been an issue such as this one, which touches all users (business, private, and government), all national and international governmental organizations, and all elements of commerce and industry (economic, legal, trade, and technology.) In part, this reflects the changing role of the World Wide Web, as well as signaling further complexities that will be encountered as the move to a massively connected world continues.

As a leader in online technology development, with a strong focus on the consumer experience, Adobe has a history of making the online experience enjoyable for consumers. As the owner of one of the largest online analytics businesses in the world, we understand the benefits of first party tracking, for first party uses, for the purpose of improving the online experience for consumers. We also believe that any interaction with consumers must be based on the principles of trust, mutual understanding, and integrity. We work to strike a proper balance – we understand that companies want to offer customers meaningful content and high-impact online interactions. Equally important, consumers want to experience the Internet in ways that speak to their unique interests. In every case, however, safeguarding consumer privacy is paramount.

## Summary of Adobe's Position

Adobe will support and participate in industry or standards initiatives that foster clear and meaningful choice regarding online tracking for purposes that are not obvious in context or commonly accepted, as described in the Federal Trade Commission's December 2010 Preliminary Staff Report. Adobe supports any discriminating "Do Not Track" mechanism that empowers, protects, and informs consumers that does not hamper innovation -- this is good for consumers and competition, and the many positive and necessary uses of data. These mechanisms should provide consumers with a clear understanding about the tracking to which they are opting-out.

The current *tracking* concern raised in the FTC’s Preliminary Staff Report relates primarily to the use of information obtained by tracking a user’s online activities for purposes that are not commonly accepted. The Report has a large focus on tracking for purposes of behaviorally targeting advertisements, but does not limit it to this use. Even the FTC, the consumer protection watchdog of the United States, does not take the position that all tracking violates a user’s privacy. Rather, the Commission recognizes – properly – that it is the use of the information obtained by the tracking technology, taking into account users’ reasonable expectations under the circumstances, that should be considered when determining whether privacy interests are implicated.

In its Preliminary Staff Report, the FTC took the position (albeit, preliminarily, pending its consideration of stakeholder comments) that *commonly accepted practices* do not require express consumer consent precisely because they are commonly accepted. Product fulfillment, fraud protection, and first party marketing are all listed within this category. So is the practice of websites collecting *information about visits and click-through rates to improve site navigation*. This falls within the preliminary set of commonly accepted practices because, just as offline retailers use consumer data to optimize their limited shelf space, websites need consumer data to optimize their sites. As such, the FTC does not believe this practice would require user consent. This form of tracking is distinguished, for example, from the unanticipated practice of selling personal information to third parties for secondary purposes unrelated to the purposes for which the data was originally collected. An industry standard solution geared towards protecting users from *unwanted tracking* should clearly define the specific type of tracking on which the solution focuses.<sup>1</sup>

Moving forward, we believe that “clear and meaningful choice” requires clear and meaningful definitions of the problem, its component parts, and its proposed solutions. Defining the problem requires understanding consumers’ reasonable expectations. Only then can we determine where the tracking related solutions are required. Some of the current tracking proposals that have been announced by various browsers address many issues, some of which may not even pose threats to privacy. It is imperative that stakeholders define the problem we are trying to solve as a first step.

After the problem has been defined, the second step should be to reach a consensus on a clear set of definitions of the component parts of the problem. Without a clear set of definitions, we will continue to provide solutions that may or may not address real privacy issues and consumer

---

<sup>1</sup> The industry standard should also strive to satisfy the five requirements set out by the FTC: (1) a Do Not Track solution should be implemented universally, i.e. one-opt out that would apply to all sites that track; (2) the solution should be easy to find, easy to understand and easy to use; (3) the user’s choice should be persistent, i.e. not deleted unless the user intended the deletion; (4) the solution should be effective and enforceable; and (5) the opt-out should apply to all defined tracking and relevant uses. As we discuss tracking – the problem and potential solutions – we need to keep in mind that the various initial solutions offered by the browser companies should gravitate to these five tenants or risk regulation.

concerns. More importantly, we risk doing harm to consumers' expectations and degrading the online experience for ordinary users. Just as privacy engenders trust, and therefore stimulates the continued growth of ecommerce, so does a positive, intuitive, engaging, and ever-improving consumer experience. Standards need to take into account both sets of reasonable end-user expectations and ensure that any solution retains equilibrium between the two. Tilting the balance too far in either direction does equal harm to the same objective: retaining an ecosystem that supports continued and increased trust and engagement online.

Adobe has a strong stake in personal privacy and user trust. Adobe's Omniture Business Unit is a leading provider of web analytic services that enables customers to capture, store, and analyze information generated by the use of their web sites to gain critical business insights into the performance and efficiency of their site, marketing and sales initiatives, and other business processes. Although the data generated by Adobe's products resides on Adobe's servers, each customer owns the data generated by the use of its site. By contract, Adobe has no right to access or use this data. In addition, Adobe does not allow use of the data for any purpose other than those of the owner (web publisher); that is, Adobe silos each customer's data for use by that customer.

Users benefit from this form of tracking. It enables streamlined paths through websites uniquely created by careful analysis of usage patterns and common needs and results in more engaging online experiences. Being able to bring the right information to the user at the right time benefits both the user and the business.

Another aspect of Adobe's business that is relevant to this discussion is its Flash technology platform. Local storage used by Flash Player (sometimes referred to as *Flash Cookies*) may be used to track users in place of cookies. It will be important for Adobe to understand the implementation of a Do Not Track solution to ensure that the user's choice is relayed to the Flash developer. It is not possible for Adobe to know how the local storage is being used by developers. It will be up to each developer to honor the user's tracking choice.

## Conclusion

Adobe fully supports measures to enable web users to have control over their privacy and their personal information. Adobe has a stake in finding suitable protections that empower consumers and build the foundations of trust that are necessary for ecommerce to continue to grow and thrive.

The ecosystem is complex. User expectations and assumptions are similarly complex. Any "fix" requires a clear articulation of the harm to be addressed and a solution narrowly tailored to address that harm. Simple solutions that prohibit all collection of data fail both prongs of this test. Assuming that all tracking is harmful, or even potentially so, is just as dangerous to the

ecosystem as assuming all tracking is benign. Addressing the assumption with a blunt instrument fails the narrowly-tailored test, and, by definition, risks collateral damage with no corresponding consumer benefit. Addressing all tracking with a single solution will confuse and frustrate users, perhaps even more so than they are frustrated now with no solution.

Adobe supports a discriminating Do Not Track solution that results from defining the problem from the perspective of consumers' expectations and defining key terms. Working together we need to identify the harm that must be addressed to foster trust and preserve the ecosystem without going so far as to cause frustration from unexpected and poor online experiences. We should focus on what consumers want and expect in terms of privacy and their online experience and tailor a solution that optimizes both.



## **Self-Tracking on the Web: Why and How**

Mathieu d'Aquin, Matthew Rowe and Enrico Motta  
Knowledge Media Institute, The Open University  
Walton Hall, Milton Keynes, MK7 6AA, UK  
{m.daquin,m.c.rowe,e.motta}@open.ac.uk

Social, professional or commercial interactions on the Web rely extensively on the exchange of private, personal information. This is already the case in the offline world where disclosing certain personal information is necessary to enable engagement with other people and organisations. However, on the Web, the circulation of such information is happening in an un-restrained, fragmented and distributed environment, making it difficult for individuals to monitor and control what is being exposed and shared about them. In other words, while personal information, interests and habits are being tracked by a large number of websites and organisations through various mechanisms and for various purposes, individual Web users are mostly unaware of the type of information they expose and that is circulated about them on the Web.

In this position paper, we argue for the need for better consideration of the activity of self-tracking - i.e., the activity of monitoring and analysing one's own behaviour regarding personal information exchange and the consequences of such behaviour on their exposure, privacy and reputation. Indeed, recently there have been growing concerns regarding the way personal information is handled by the organisations collecting it, and how such information could be used to the disadvantage of Web users. Amongst the most cited issues are identity theft, lateral surveillance and data aggregation to the benefit of commercial companies or for malevolent activities. However, as our preliminary experiments have shown [d'Aquin et al., 2010a], the inherent complexity and fragmentation of the flow of personal information on the Web makes it impossible for an individual Web user to monitor, make sense of and act on his/her own exposure without appropriate technological support. In contrast with such complexity, the tools currently available to Web users are extremely limited. More and more users would simply use popular Web search engines to check websites where their name appears, however with all the noise and ambiguities that such a method introduces [Madden and Smith, 2010] the effectiveness and success of such an approach is limited.

The requirement to achieve effective self-tracking appears with respect to such issues, in an environment as complex as the Web. It can be seen as a specific approach to lifelogging (called Web lifelogging in [d'Aquin et al., 2010a]) focusing on Web interactions, with the purpose of providing sufficient data to achieve appropriate levels of personal information management [Jones and Teevan, 2007], personal reputation management, and of course, privacy.

While appearing as such a crucial need, support for self-tracking on the Web has remained mostly unexplored, apart from isolated initiatives and tools focusing on specific issues. Here, we review such initiatives and tools with the aim to identify a path towards a more principled and comprehensive approach to self-tracking. We distinguish two major trends in existing work: tracking one's own behaviour in terms of Web interactions and exchange of personal information, and tracking the appearance of one's personal information on the Web.

### **Tracking one's own Web interactions, traffic, behaviour**

Research, as well as many commercial developments, have until now mostly been dedicated to logging user visits to websites, in order to provide valuable information to website owners in the form of patterns of interactions. However, tools such as Google Web History<sup>1</sup> can be

---

<sup>1</sup>[www.google.com/psearch](http://www.google.com/psearch)

used to record different aspects of Web activities, as long as they are done in the scope of what can be perceived by Google systems. Such an approach provides an interesting starting point to collecting information regarding one's own behaviour online, but has obvious limitations, including the lack of comprehensiveness and control over what is being collected, as well as the need to go through a third party (Google).

The perceived gap in the ability of users to take ownership of their own Web activity data has led to the emergence of the notion of attention data<sup>2</sup>, with tools such as the Attention Recorder<sup>3</sup> developed explicitly to provide the user with ways to track their Web activity, as carried out through a browser. The idea here is that the user can claim back their own activity data, so that they can be shared and traded in their own terms. Technically, tools such as the Attention Recorder still need to gain maturity, to be able to cover the wide variety of sources of activity (attention) data on the Web, and to provide appropriate support for the user to truly exploit the collected information.

In [d'Aquin et al, 2010a] we experimented with the idea of a complete, unrestricted 'self-monitoring' of personal, online activities, in a process comparable to the idea of *lifelogging* [O'Hara et al, 2009]. Even in relatively small settings, such an approach provides rich data about the user's behaviour [d'Aquin et al., 2010a], using a "local Web proxy" to obtain Giga Bytes of information about a single user's Web activities within the scope a 2.5 months. Specific analyses of the data collected revealed promising potential for such an approach. Simple geographical mappings of the requests from the user shown expected patterns, with most of the activities concentrating in Europe and North America, but also helped identifying anomalies (e.g., a small number of requests to Nigeria) that could be explored further based on the collected data. Looking at other indicators, such as the number of requests to different websites, the quantity of information transferred to these sites, and the user agents used in these transactions also demonstrated the extent to which activities and exchanges on the Web are "implicit", i.e., realized without being explicitly triggered by the user. More sophisticated analyses based on the keywords used to query search engines showed how such simple information can be used to build a profile of the interests of the user, according to a particular view which might not be the one he or she is prepared to expose. There is indeed a generalized discrepancy between the user's view of his/her own behaviour on the Web, and the reality of this behaviour as it can be perceived through self-tracking. To illustrate this point, in [d'Aquin et al., 2010b], we devised a model of the observed trust in websites and criticality of pieces of personal information, which is derived from the traces of activities collected for an individual user. The idea is that, through exposing users to such an abstract view of their own behaviour online, they can make emerge such discrepancies, leading to a better understanding and an improved awareness of the potential consequences of exposing personal information.

The idea of "logging" one's own Web activities is still in an early stage and the potential for analysis of such an approach remains mostly unexplored. In other terms, Web lifelogging faces similar challenges to other forms of lifelogging, including the need for mechanisms to abstract and interpret the obtained low-level raw data into something exploitable by the user [d'Aquin, 2010].

### **Tracking one's references on the Web**

Besides tracking one's own behaviour, a key to self-tracking is the ability to monitor what information about an individual has been made visible on the Web, possibly without the user's consent. Web presence is an important aspect of business and reputation for the majority of Web users. Inflammatory content or misleading information can have dire consequences for the individual that it describes, for instance, [Andrejevic, 2005] cites

---

<sup>2</sup> see e.g., [http://majestic.typepad.com/seth/2005/10/atx\\_the\\_attenti.html](http://majestic.typepad.com/seth/2005/10/atx_the_attenti.html)

<sup>3</sup> <http://addons.mozilla.org/en-US/firefox/addon/3569/>

examples of employers ‘vetting’ prospective employees by searching the Web for information about them. The recent Javelin report<sup>4</sup> describes the 2010 identify fraud statistics collected from US companies, showing an overall reduction in the number of cases, while the mean economic cost of such cases has risen – indicating a move towards targeting selective individuals. Individual web users must be informed where their personal information resides on the Web, so that the correct action may then be taken – i.e. applying for the information to be removed if it has been placed there without consent, or altering the visibility settings of the profile if the user has intentionally placed it there.

The sheer scale of the Web however makes manually finding web references largely infeasible. Automatic methods and third party services therefore provide a viable solution to overcoming such tasks. Identifying web citations is a single-person disambiguation task: given a collection of Web pages, all of which contain a specific person’s name, the goal is to disambiguate those pages which refer to the individual of interest. Our experience [Rowe and Ciravegna, 2010] shows that an efficient approach is to use a combination of supervised classification models with a semi-supervised framework. A common issue when applying such methods is obtaining initial seed data to start the identification process. For instance, we may only know a few web references for the individual, the information from which we can use as seed data describing the person. Using such a framework, therefore, allows information to be learnt in an on-going process as more web references are found and the information within those web references put to use.

The extraction of such information also poses a problem. The messiness of information provided on the Web, given the heterogeneous nature of HTML and the lack of conformance to web standards, makes it hard for machines to parse web pages for personal information. Techniques are therefore required that can effectively extract personal information from the Web at high-levels of accuracy. Furthermore, as mentioned previously, information published on the Web about an individual may damage the person’s reputation if it is negative or describes the individual in a bad way. Sentiment analysis techniques are therefore required which can assess the sentiment, or feeling, towards the person in the web page, enabling reputation assessment in an automated fashion at a large-scale.

Several companies have tackled the above issues, for example SentiMetrics<sup>5</sup> use social media sources to calculate the sentiment towards a given person based on available information, and Trackur<sup>6</sup> and Visible Technologies<sup>7</sup> also monitor social media sites for references to a person. Garlik’s Data Patrol<sup>8</sup> service assesses the risk of an individual to identity theft, based on the presence of their sensitive information on the Web. Identity Guard<sup>9</sup> provides a service that monitors a person’s information distributed across the Web, and alerts the individual when the exposure of his/her information could have a detrimental effect.

While such existing services tackle the individual aspects of web exposure, a single unifying approach is currently lacking that informs the web user where their personal information resides on the Web, the sentiment that such references have, and ultimately how the visibility of such information could effect the person. Therefore a core, unsolved challenge is to integrate and relate all these different pieces of information, to understand and interpret them in a context which takes into account the user’s identity, activities and own perception of his or her exposure.

---

<sup>4</sup> <http://www.idsafety.net/report.php>

<sup>5</sup> <http://www.sentimentmetrics.com/>

<sup>6</sup> <http://www.trackur.com/>

<sup>7</sup> <http://www.visibletechnologies.com/>

<sup>8</sup> <http://www.garlik.com/dpindividuals.php>

<sup>9</sup> <http://www.identityguard.com>

## Conclusion

More and more personal information is being shared, exchanged and exposed by Web users everyday, mostly without their consent and awareness. A lot of efforts and attention is currently being given to the way online organizations might track this information, to their own benefit, and potentially, to the detriment of the users. Here, we discussed initial tools and techniques towards taking the inverse perspective: helping Web users tracking and monitoring their own personal information online, to their own benefit.

As our initial experiments have shown, achieving such a process of self-tracking can be very revealing to Web users, helping them reaching a better awareness of their own online behaviour, and a better understanding of the possible consequences of such behaviour on the exposure of their personal information. Such an approach appears to be crucially needed as the Web evolves to both a global information marketplace, and a major medium for all sorts of social interactions online. However, the tools and technologies currently available to carry out self-tracking on the Web are inadequate, to the point that many Web users would resort to using a Web search engine to check where their name appears [Madden and Smith, 2010].

We therefore argue that a more principled and comprehensive study of the activity of self-tracking on the Web and of the technological requirements for such an activity to take place should be conducted. This requires for both the social and conceptual models of the way personal information is exchanged on the Web to be related to the technological protocols that are used as mediums for instantiating these models. From a more concrete point of view, we believe that a new set of tools are to be created that will support users in monitoring their own activity on the Web, tracking the appearance of their personal information online, and interpreting this information in terms of behaviour, reputation and privacy risks. A positive effect of the availability of such tools is not only to provide individuals with better control over the exposure of their information, but also to support a generic understanding of the global mechanisms underlying such circulation of personal information on the Web.

## References

- [Andrejevic, 2005] M. Andrejevic (2005) The work of watching one another: Lateral surveillance, risk and governance. *Surveillance and Society*, 2 (4):479–497, 2005.
- [d'Aquin, 2010] M. d'Aquin, (2010) Making Sense of Users' Web Activity, Personal Semantic Data, PSD (keynote) at EKAW 2010.
- [d'Aquin et al, 2010a] d'Aquin, M., Elahi, S. and Motta, E. (2010) Personal Monitoring of Web Information Exchange: Towards Web Lifelogging, Poster at Web Science 2010 Proceedings of the WebSci10: Extending the Frontiers of Society On-Line
- [d'Aquin et al, 2010b] d'Aquin, M., Elahi, S. and Motta, E. (2010) Semantic Monitoring of Personal Web Activity to Support the Management of Trust and Privacy, Workshop: SPOT 2010 - 2nd Workshop on Trust and Privacy on the Social and Semantic Web at ESWC 2010
- [Jones and Teevan, 2007] W. Jones and J. Teevan (editors) (2007), *Personal Information Management*, University of Washington Press
- [Maden and Smith, 2010] M. Madden, A. Smith (2010), *Reputation Management and Social Media*, Report from the PewResearchCenter  
(<http://www.pewinternet.org/Reports/2010/Reputation-Management.aspx>)
- [O'Hara et al, 2009] K. O'Hara, M. Tuffield and N. Shadbolt (2009) Lifelogging: Privacy and Empowerment with Memories for Life. *Identity in the Information Society*, 1 (2).
- [Rowe & Ciravegna, 2010] M. Rowe and F. Ciravegna, (2010) Disambiguating Identity Web References using Web 2.0 Data and Semantics. *The Journal of Web Semantics*.

# Position Paper for W3C Workshop on Web Tracking and User Privacy

Li Li, Wu Chou

Avaya Labs Research, Avaya, USA

Web privacy and tracking protection technologies will have impact to enterprise applications, as more and more enterprise applications are based on the Web infrastructure and technologies. In a typical scenario, a user on the enterprise Intranet would use the same Web browser to access both internal and external Web sites. It may be unrealistic or inconvenient to ask a user to change the Web browser settings on the fly based on which site to visit. Therefore, settings of the Web browser are likely to be shared internally and externally.

For this reason, a global binary “Do Not Track” option applicable to all Web sites may not be suitable for enterprise applications, as enterprise needs to track the use of the information on the Web for reasons such as security and service quality. Such a binary option in the browser may disable an internal Web site that uses Web tracking technologies to improve organizational productivity. For example, an internal Web site may embed a Web beacon from an enterprise tracking service to collect an employee’s online activities, while at the same time, the employee does not want to be tracked by any external advertisement site. It is possible that this tracking service is deployed in a trusted third party domain outside the enterprise Intranet or in a hosted cloud. In any case, when the employee turns on “Do Not Track” in the Web browser, both the unwanted and wanted tracking will be blocked.

We hope Web browser vendors can adopt a privacy framework that allows for finer-grained tracking control, such that enterprise privacy policies along with personal preferences can be both incorporated in a Web browser. We think the Tracking Protection List (TPL) introduced by Microsoft IE9 [1] is a good starting point. In addition, we think it is useful to have two levels of tracking controls, one for enterprise policies, and one for personal preferences.

It is possible to enforce enterprise policies at a HTTP proxy shared by the browsers. However, many Web browsers do not use a proxy at all for performance reasons. A transparent proxy can enforce enterprise tracking policies without any browser configuration.

However, in reality, it needs to enforce the tracking policies in the presence of personal preferences, and personal preference may block the enterprise tracking policies if they are deployed on the proxy, including transparent proxy. For example, if the personal preferences block a tracking site xyz.com, but tracking policies deployed at the proxy allow it, then the site will still be blocked if the browser does not send any request to the proxy. For these reasons, a browser-based configuration is a more attractive solution.

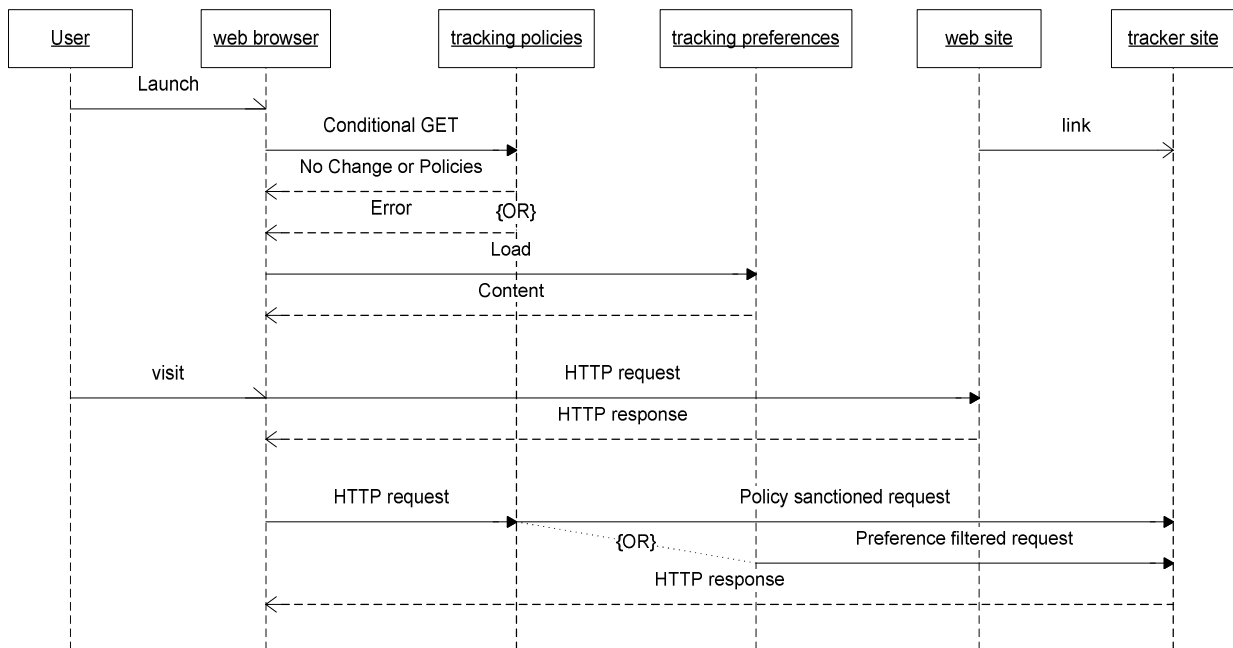
To realize such a solution in browser, the enterprise tracking policies, whenever available, will have precedence over the personal ones. Combined with “Do Not Track,” this leads to a layered tracking protection approach against a tracking site as follows:

- 1) if it is in enterprise policies, then use the matched policy; otherwise,
- 2) if it is in personal preferences, then use the matched preference; otherwise,
- 3) use the “Do Not Track” option.

The enterprise tracking policies are specified in a XML file located in an internal enterprise Web server only accessible on the enterprise network, possibly through VPN. This file is managed by authorized administrators and can be consulted by a Web browser using HTTP GET to that URL. The personal privacy preferences are managed by a user through the Web browser interface.

As notebook computers and smart phones can move in and out of an enterprise network, the enterprise tracking policies should be activated or deactivated accordingly. This process can be automated by a Web browser sending a Conditional HTTP GET to the tracking policies URL at the browser startup time. If this Conditional HTTP GET request succeeds, then the Web browser is inside the enterprise and on the enterprise Intranet. Otherwise, the enterprise tracking policies are not consulted and only personal preferences are used.

The flow diagram of this approach is depicted below that illustrates the interactions between each related components, where “Do Not Track” option is treated as part of the tracking preferences.



The advantage of this approach is that the applicability of tracking policies is managed automatically by a browser during the browser startup time. A disadvantage is that it may increase browser startup time, as it needs to do a Conditional HTTP GET for enterprise tracking policies. But this occurs only once at the time when the browser starts up.

To ensure the interoperability between Web browsers and tracking controls (XML files), tracking controls should be standardized. The standard should enable the switch of Web browser while

maintaining the tracking preferences systematically without ad-hoc translations that might introduce inconsistency. As an advantage of the described approach, there is no need to import enterprise tracking policies except passing the enterprise tracking policies URL, as enterprise tracking policies are automatically loaded when the browser starts up inside the enterprise.

[1] <http://blogs.msdn.com/b/ie/archive/2010/12/07/ie9-and-privacy-introducing-tracking-protection-v8.aspx>

# SUMMARY COMPARISON OF UNIVERSAL OPT-OUT MECHANISMS FOR WEB TRACKING

Alissa Cooper / March 25, 2011

This table summarizes and builds upon “Overview of Universal Opt-Out Mechanisms for Web Tracking,” an Internet draft recently submitted to the IETF, available at <http://tools.ietf.org/html/draft-cooper-web-tracking-opt-outs-00>. Consult the draft for a fuller discussion and comparison of web tracking opt-out mechanisms.

	Domain/request blocking	Do Not Track HTTP header	Do Not Track DOM property
<b>Universality</b> <i>Does it work across domains/apps/entities?</i>	Relies on extent to which domains/resources that conduct tracking are included on block lists	Can be sent with every HTTP request	Can be made accessible to all sites that access the DOM
<b>Effectiveness</b> <i>How well does it prevent tracking?</i>	Prevents tracking altogether from domains/resources on block lists	Relies on how tracking is defined and extent to which tracking entities honor the header  May require enforcement or intervention from governmental privacy authorities	Relies on how tracking is defined and extent to which tracking entities honor the property  May require enforcement or intervention from governmental privacy authorities
<b>Comprehensiveness</b> <i>Does it work for different tracking technologies?</i>	Applies to tracking via any mechanism that originates with a web server request (cookies, other HTTP headers, script-based techniques, etc.)	Can be defined to apply to tracking via any mechanism employed by HTTP servers	Can be defined to apply to tracking via any mechanism employed by client-side documents
<b>Simplicity</b> <i>How easy is it to use?</i>	Requires block list to be installed and kept up-to-date	Can be offered via simple binary choice with possibilities for more granular choices	Can be offered via simple binary choice  Offering granular choices is more complicated because DOM is shared across domains
<b>Continuity with web functionality</b> <i>How does it impact existing web sites and applications?</i>	Prevents content delivery from domains used for both tracking and content serving  Domain operators could seek to avoid being blocked by switching domains or requiring users to disable block lists to access content	Does not directly interfere with existing functionality  Sites that detect the header may prevent users from accessing content or may request that users turn it off before access is granted	Does not directly interfere with existing functionality  Scripts that detect the property may prevent users from accessing content or may request that users turn it off before access is granted
<b>Standardization</b> <i>What components could or should be standardized?</i>	Block list format and processing rules	Syntax, semantics and usage	Syntax, semantics and usage





**ELECTRONIC PRIVACY INFORMATION CENTER**

---

Statement for the Record of  
The Electronic Privacy Information Center (EPIC)

Marc Rotenberg, EPIC President  
Sharon Gcott Nissim, EPIC Consumer Protection Fellow

Hearing on  
“Do Not Track Legislation: Is Now the Right Time?”

Before the  
Committee on Energy and Commerce;  
Subcommittee on Commerce, Trade and Consumer Protection;  
U.S. House of Representatives

December 2, 2010  
2123 Rayburn House Office Building  
Washington, D.C.

Mr. Chairman, Members of the Committee, this statement was prepared for the hearing “Do Not Track Legislation: Is Now the Right Time?” held on December 2, 2010 before the House Committee on Energy and Commerce Subcommittee on Commerce, Trade, and Consumer Protection. We ask that it be included in the hearing record.

The Electronic Privacy Information Center (EPIC) is a non-partisan public interest research organization established in 1994 to focus public attention on emerging privacy and civil liberties issues. EPIC has long focused on the impact of emerging technologies on privacy. And I was directly involved in the development of the Telephone Consumer Protection Act of 1991 and the Do Not Call program that followed, which established a meaningful and effective way for consumers to opt-out of telemarketing calls.

EPIC supports the Committee’s examination of Do Not Track proposals. It is important to recognize that as the Internet has expanded, so have the invasions of consumer privacy, in the form of data collection and behavioral targeted advertising. EPIC recommends that the Committee evaluate the Do Not Track proposal in light of the lessons from past efforts to safeguard consumers from unwanted advertising and marketing.

## **I. The History of the Telephone Consumer Protection Act (TCPA) and Do Not Call List**

In this current debate over a Do Not Track system for the Internet, it is helpful to look back and examine previous debates over the Telephone Consumer Protection Act and the Do Not Call List. While any future Do Not Track mechanism may look different from the Do Not Call registry, many of the issues encountered then are still relevant now.

### *A. The Telephone Consumer Protection Act (TCPA)*

The Telephone Consumer Protection Act, 42 U.S.C. § 227, was passed in 1991. This Act amended the Communications Act of 1934 to prohibit automated and prerecorded telephone calls to the home, as well as the sending of unsolicited fax messages.<sup>1</sup> The Act directed the Federal Communications Commission (FCC) to initiate a rulemaking proceeding concerning the need to protect peoples' privacy rights to avoid receiving telephone solicitations they do not want, including the possibility of establishing a single national database compiling a list of those residents who object to such phone calls.<sup>2</sup> The Act allowed states to bring civil suits to enforce the law,<sup>3</sup> but gave exclusive jurisdiction over these actions to federal district courts,<sup>4</sup> and also provided for a private right of action.<sup>5</sup>

---

<sup>1</sup> 42 U.S.C. § 227.

<sup>2</sup> 42 U.S.C. § 227 (c)(1)(A).

<sup>3</sup> 42 U.S.C. § 227 (f)(1).

<sup>4</sup> 42 U.S.C. § 227 (f)(2).

<sup>5</sup> 42 U.S.C. § 227 (c)(5).

## *B. The Creation of the Do Not Call Registry*

The FCC, as directed by the TCPA, initiated a rulemaking on the idea of a Do Not Call registry and other related matters.<sup>6</sup> EPIC, along with ten other advocacy groups, submitted comments urging the creation of a telemarketing "do not call" registry.<sup>7</sup> The comments identified the public's frustration with the "intrusion into the privacy of the home," of unwanted telephone solicitations, and described how difficult it was under the current rules for individuals to prevent these type of calls, especially in light of changing technologies.<sup>8</sup> Additionally, the comments laid out the legal reasoning as to why the FCC's proposed regulations were consistent with First Amendment principles.<sup>9</sup>

The EPIC comments also pointed out, however, that an opt-in system requiring express consent from individuals before telemarketers could initiate sales calls would be preferable to the opt-out regime that a Do Not Call registry imposes. "An opt-in framework," the comments explained, "would better protect individuals' rights and is consistent with most United States privacy law."<sup>10</sup> The EPIC comments argued further that opt-in is more effective "because it encourages companies to explain the benefits of information sharing, and to eliminate barriers to exercising choice . . . [e]xperience with opt-out has shown that companies tend to obfuscate the process of exercising choice, or that exemptions are created to make opt-out impossible."<sup>11</sup>

The FTC also proposed the Telemarketing Sales Rule (TSR),<sup>12</sup> which included a do not call list, and received similar favorable comments from EPIC and other groups in response.<sup>13</sup> These new FTC regulations required telemarketers to transmit caller ID information, establish new rules for the use of preacquired account number information, and prohibit "abandoned" calls.<sup>14</sup>

---

<sup>6</sup> FCC Notice of Proposed Rulemaking on the TCPA, Oct. 8, 2002, *available at* [http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=2002\\_register&docid=02-25569-filed](http://frwebgate.access.gpo.gov/cgi-bin/getdoc.cgi?dbname=2002_register&docid=02-25569-filed).

<sup>7</sup> Comments of EPIC, et al. before the FCC, in the matter of "Rules and Regulations Implementing the Telephone Consumer Protection Act of 1991," Dec. 9, 2002, *available at* <http://epic.org/privacy/telemarketing/tcpacomments.html>.

<sup>8</sup> *Id.*

<sup>9</sup> *Id.*

<sup>10</sup> *Id.*

<sup>11</sup> *Id.*

<sup>12</sup> Federal Trade Commission, Telemarketing Sales Rule, 16 CFR Part 310, *available at* <http://www.ftc.gov/os/2002/12/tsrfinalrule.pdf>.

<sup>13</sup> Comments of EPIC et al, before the Federal Trade Commission, in the matter of Telemarketing Rulemaking – Comment, April 10, 2002, *available at* <http://epic.org/privacy/telemarketing/tsrcomments.html>.

<sup>14</sup> Federal Trade Commission, Telemarketing Sales Rule, 16 CFR Part 310, *available at* <http://www.ftc.gov/os/2002/12/tsrfinalrule.pdf>.

In March 2003, Congress passed legislation allowing the FTC to operate a national Do Not Call List.<sup>15</sup> This legislation approved the levying of fees on the telemarketing industry in order to fund this program.<sup>16</sup> In June of 2003, the National Do Not Call Registry opened for enrollment and registration exceeded 10 million on the first day.<sup>17</sup> Registry enforcement is coordinated between the FCC and the FTC according to a memorandum of understanding.<sup>18</sup> As of October 2003, 53.7 million numbers were registered on the Do Not Call list and consumers had filed 15,000 complaints against telemarketers who did attempt to call them.<sup>19</sup>

Originally the FTC adopted a five-year re-registration mechanism for the Do Not Call list to ensure it was accurate.<sup>20</sup> However, the FTC has successfully used a scrubbing program to purge the Registry of disconnected and reassigned numbers each month.<sup>21</sup> This program, along with the increased use of cell phones and the popularity of telephone number portability, made the re-registration procedure less necessary than it had been when it was adopted.<sup>22</sup> On October 23, 2007, the FTC testified before Congress that "it will not drop any telephone numbers from the Do Not Call Registry based on the five-year expiration period pending final Congressional or agency action on whether to make registration permanent."<sup>23</sup>

### *C. Legal Challenges to Do Not Call*

Industry groups immediately responded to the creation of the Do Not Call registry by filing lawsuits. Several lawsuits were filed, arguing that the Do Not Call registry was unconstitutional under the First Amendment because it did not protect corporate telemarketers' "commercial speech" and the exclusion of non-commercial charitable

---

<sup>15</sup> "Do Not Call Implementation Act," Public Law 108-10.

<sup>16</sup> *Id.*

<sup>17</sup> Federal Trade Commission, June 17, 2003, "Do Not Call Registrations Exceed 10 Million," available at <http://www.ftc.gov/opa/2003/06/dncregistration.shtm>.

<sup>18</sup> See FTC Annual Report to Congress, FY 2003 and 2004, "Pursuant to the Do Not Call Implementation Act on Implementation of the National Do Not Call Registry," Appendix – FTC-FCC Memorandum of Understanding on Telemarketing Enforcement.

<sup>19</sup> FTC, "Consumers on Do Not Call Registry File Over 15,000 Complaints Against Telemarketers," Press Release, October 16, 2003, available at <http://www.ftc.gov/opa/2003/10/dnccomplaints.shtm>.

<sup>20</sup> See generally, EPIC: Do Not Call, available at <http://epic.org/privacy/telemarketing/dnc/>.

<sup>21</sup> *Id.*

<sup>22</sup> *Id.*

<sup>23</sup> Statement of the Federal Trade Commission, "Enhancing FTC Consumer Protection in Financial Dealings, with Telemarketers, and on the Internet," before the Committee on Energy and Commerce, Subcommittee on Commerce, Trade, and Consumer Protection, U.S. House of Representatives, Washington, D.C, Oct. 23, 2007, available at <http://www.ftc.gov/os/testimony/071023ReDoNotCallRuleEnforcementHouseP034412.pdf>.

organizations from the registry amounted to a "content-based" speech restriction.<sup>24</sup> The suits also charged that the FTC did not have authority to enact these rules.<sup>25</sup>

In February 2004 in a consolidated appeal of these suits, the U.S. Court of Appeals for the Tenth Circuit upheld the FTC's Do Not Call Registry.<sup>26</sup> The Court held that the Do Not Call registry did not violate the First Amendment, the registry is a reasonable restriction on commercial speech,<sup>27</sup> and "commercial calls were more intrusive and posed a greater danger of customer abuse."<sup>28</sup> The Court also found that the FTC had the authority to create and operate the list, and could levy fees on telemarketers for its operation.<sup>29</sup>

## II. Online Advertising and Privacy

This section presents an overview of the current problems in online tracking and targeted advertising. Marketing has come a long way from telephones, and online advertisers use a variety of web-based tactics to track consumers' online behavior and target ads based on that behavior.

### A. Data Collection

There is a giant chasm between the type of tracking that companies are engaged in on the web and what people know or think is occurring. The general public has very little idea that every second they are on the Internet, their behavior is being tracked and used to create a "profile" which is then sold to companies on "stock-market-like" exchanges.<sup>30</sup> According to a Wall Street Journal study, the nation's top five websites installed an average of 64 pieces of tracking technology onto the computers of visitors, usually without warning, for a total of 3,180 tracking files. A dozen sites installed more than a hundred.<sup>31</sup> Two-thirds of those files installed by 131 companies that are in the tracking and online consumer profiling business.<sup>32</sup>

Online tracking is no longer limited to the installation of the traditional "cookies" that record websites a user visits. Now, new tools can track in real time the data people are accessing or browsing on a web page and combine that with data about that user's

---

<sup>24</sup> See *Mainstream Marketing Services, Inc. v. FTC*, 283 F.Supp.2d 1151 (D. Colo. 2003); *U.S. Security v. FTC*, 282 F.Supp.2d 1285 (W.D. Okla. 2003).

<sup>25</sup> *Id.*

<sup>26</sup> *Mainstream Marketing Services, Inc., et al. v. Federal Trade Commission, et al.*, 358 F.3d 1228 (10<sup>th</sup> Cir. 2004), available at <http://www.epic.org/privacy/telemarketing/03-1429.pdf>.

<sup>27</sup> *Id.* at 1237-39 (finding substantial government interest in "1) protecting the privacy of individuals in their homes, and 2) protecting consumers against the risk of fraudulent and abusive solicitation," and a reasonable fit between the rules and these interests).

<sup>28</sup> *Id.* at 1233.

<sup>29</sup> *Id.* at 1246-50.

<sup>30</sup> Julia Angwin, "The Web's New Gold Mine: Your Secrets," *What They Know Series*, THE WALL STREET JOURNAL, July 30, 2010.

<sup>31</sup> *Id.*

<sup>32</sup> *Id.*

location, income, hobbies, and even medical problems.<sup>33</sup> These new tools include flash cookies and beacons. Flash cookies can be used to re-install cookies that a user has deleted, and beacons can track everything a user does on a web page including what the user types and where the mouse is being moved.<sup>34</sup>

Advertisers are no longer limited to buying an ad on a targeted website because they are now paying to "follow people around the Internet, wherever they go, with highly specific marketing messages."<sup>35</sup> Companies then use this information to decide what credit-card offers or product pricing to show people, potentially leading to price discrimination.<sup>36</sup>

### *B. Privacy Issues*

This type of data collection violates several Fair Information Practices (FIPs).<sup>37</sup> These online tracking companies have no transparency – so there is no way for a user to access the data being collected about him or her, or correct any inaccuracies. And even if users were to somehow be able to find out what information was being collected, they have no control over what the data collecting companies subsequently do with that information.

According to the Consumer Federation of America and Consumers Union, "there is a fundamental mismatch between the technologies of tracking and targeting and consumers' ability to exercise informed judgment and control over their personal data."<sup>38</sup> The information being collected online is not information that consumers voluntarily share with these tracking companies or online advertising businesses. There are no regulations or limits on what can be collected.

Very sensitive information is often collected, including health and financial data. One company, Healthline, lets advertisers track people with bipolar disorder, overactive bladder, or anxiety – producing ads related to those conditions targeted at specific people.<sup>39</sup> Advertisers collect, use, and sell social security numbers, financial account numbers, and information about sexual behavior and sexual orientation with no controls or limits.<sup>40</sup>

---

<sup>33</sup> *Id.*

<sup>34</sup> *Id.*

<sup>35</sup> *Id.*

<sup>36</sup> Emily Steel and Julia Angwin, "On the Web's Cutting Edge, Anonymity in Name Only," What They Know Series, THE WALL STREET JOURNAL, August 4, 2010.

<sup>37</sup> Code of Fair Information Practices, *available at* [http://epic.org/privacy/consumer/code\\_fair\\_info.html](http://epic.org/privacy/consumer/code_fair_info.html).

<sup>38</sup> CFA and CU comments to the FTC concerning the Proposed Online Behavioral Advertising Self-Regulatory Principles, April 11, 2008, *available at* [www.ftc.gov/os/comments/behavioraladprinciples/080411cfacu.pdf](http://www.ftc.gov/os/comments/behavioraladprinciples/080411cfacu.pdf).

<sup>39</sup> Angwin, *supra* note 30.

<sup>40</sup> CFA and CU comments, *supra* note 38 at 4.

Another consequence of online data collection is the possibility that all these "anonymized" pieces of data could actually be used to identify a person. In the Wall Street Journal, a researcher described how all that is needed to "de-anonymize" data is 33 "bits" of information (some more valuable than others) – and one exemplar website transmitted 26.5 bits of information about a user – enough to narrow the user down to one of just 64 people in the world.<sup>41</sup>

### *C. Lack of Action*

So far, online advertising and behavioral tracking companies have been allowed to operate unchecked. The FTC has relied on "notice and choice" and self-regulation as their tools of choice. But neither of these is effective at protecting consumers' privacy. Privacy policies and notices do not work; less than one percent of consumers read these statements, and even those who do read them do not generally assume that their information is shared with others or combined with information from other sources to form a profile.<sup>42</sup>

And self-regulation certainly is not the answer. The companies engaged in these tactics will not voluntarily decide to curtail them – not when it means less revenue. When given the chance, companies tend to obfuscate the process of exercising choice, or ensure that exemptions are created to make meaningful choice or opt-out impossible.<sup>43</sup> A group called the Network Advertising Initiative (NAI), composed of 11 advertiser members, says the industry polices itself and people can download an opt-out cookie.<sup>44</sup> However, not all behavioral advertising companies join this initiative, and, more importantly, the opt-out process is technically difficult and requires a different download for each advertising company from which a user wishes to opt-out.<sup>45</sup> In fact, as EPIC has earlier noted, the NAI "opt-out cookie" is counterintuitive because it requires consumers who are seeking to protect their privacy to download and retain a tracking technique when the better practice would be to simply delete all advertising related cookies.

"If you look back at the Do Not Call list it was at one time managed by industry," stated Pam Dixon, director of the World Privacy Forum.<sup>46</sup> "The industry has had seven years to prove they can manage online opt-outs. It is time to move toward something structured like the Do Not Call List to address the problems we are seeing and have now

---

<sup>41</sup> Steel and Angwin, *supra* note 36. ("bits" include income level, education, geographic location, zip code, birthdate, etc.)

<sup>42</sup> *Id.*

<sup>43</sup> Comments of EPIC, et al. before the FCC, *supra* note 7 at 4.

<sup>44</sup> See "Opt-Out of Behavioral Advertising," Network Advertising Initiative, *available at* [http://www.networkadvertising.org/managing/opt\\_out.asp](http://www.networkadvertising.org/managing/opt_out.asp).

<sup>45</sup> Catherine Rampell, "'Do Not Track' Registry Proposed for Web Use: Online Behavior Used to Tailor Ads," THE WASHINGTON POST, November 1, 2007.

<sup>46</sup> Ryan Singel, "Privacy Groups Asks for Online 'Do Not Track' List," Wired, Oct. 31, 2007, *available at* [http://www.wired.com/politics/onlinerights/news/2007/10/do\\_not\\_track](http://www.wired.com/politics/onlinerights/news/2007/10/do_not_track)

seen for seven years."<sup>47</sup> In other words, self-regulation has not worked in other consumer protection areas, and there is no reason to believe that it would work here.

EPIC believes that key to an effective Do Not Track initiative must include the adoption of legislation that makes a consumer's decision to opt out of tracking enforceable, persistent, transparent, and simple.

### III. Do-Not-Track Proposals

There are several strategies for implementing a Do Not Track system. Earlier proposals focused on registries akin to the Do Not Call list. The most recent proposals head in a different, and possibly more effective, direction.

#### A. User-Registry Approach

This approach would allow individual users to register for a do-not-track list with some unique identifier, presumably their IP address. This approach has several significant drawbacks. First, there really are no "universally recognized user identifiers" being used on the web.<sup>48</sup> "By mandating a global, robust identifier," the *33 bits* blog explains, "a user registry would in one sense *exacerbate* the very problem it attempts to solve."<sup>49</sup> This approach would also not allow a user to change do not track settings from site to site.<sup>50</sup>

Second, if IP addresses were used as the identifier, new problems emerge. IP addresses are often dynamic, and several devices can share the same IP address.<sup>51</sup> Moving to static IP addresses to enforce a Do-Not-Track system would ironically make it easier to track the activities of Internet users since the fixed IP would now operate as an "Internet SSN," and become a de facto identifier for a lot of user activity. If the registry is somehow cookie-based, then it would apply only to the browser and not the individual using it and users would have to register all their computers.<sup>52</sup>

---

<sup>47</sup> *Id.*

<sup>48</sup> "'Do Not Track' Explained," September 20, 2010, 33 Bits of Entropy, *available at* <http://33bits.org/2010/09/20/do-not-track-explained/>.

<sup>49</sup> *Id.*

<sup>50</sup> *Id.*

<sup>51</sup> Harlan Yu, "Do Not Track: Not as Simple as It Sounds," CircleID: Internet Infrastructure, Aug. 10, 2010, *available at* [http://www.circleid.com/posts/do\\_not\\_track\\_not\\_as\\_simple\\_as\\_it\\_sounds/](http://www.circleid.com/posts/do_not_track_not_as_simple_as_it_sounds/).

<sup>52</sup> Marc Roth, "The Do Not Track List and the Law of Unintended Consequences," E-COMMERCE TIMES, Oct. 16, 2010, *available at* <http://www.ecommercetimes.com/story/71048.html?wlc=1291046770>.



### *B. Domain-Registry Approach*

This approach requires advertisers that track online behavior to report what servers or domains they use to do their tracking to some authority such as the FTC.<sup>53</sup> Users would then have to download a plug-in for their browsers that would block the domains on the centralized list.<sup>54</sup> The problems with this approach are: 1) the centralization would be difficult to accomplish; 2) blocking tracking domains might block all advertisements (because showing an ad on a website necessitates contacting the hosting server); and 3) consumers must be vigilant in making sure the tracking domain list is updated.<sup>55</sup>

### *C. Current Browser-Header Approach*

This most recent idea, proposed by researchers at Stanford, is simpler and easier to execute than either of the previous approaches. In this approach, a user's browser sends a signal to a website that the user wants to opt-out of being tracked. It does so using an HTTP "header."<sup>56</sup> "Whenever a web browser requests content or sends data using HTTP, the protocol that underlies the web, it can optionally include extra information, called a 'header,'" explain the Stanford researchers.<sup>57</sup>

This mechanism "employs a decentralized design; it thus avoids the substantial technical and privacy challenges inherent to compiling, updating, and sharing a comprehensive registry of tracking services or web users."<sup>58</sup> Jonathan Mayer, one of the principal Stanford researchers, stated that while it operates differently, the Do Not Track registry, "much like the popular Do Not Call registry . . . provides users with a single, persistent setting to opt out of web tracking."<sup>59</sup>

Yet, in order to be effective, advertising companies will have to actually "listen" to this do not track signal being sent from users' browsers. According to the Stanford researchers, there are a variety of ways that this could be enforced, including self-regulation, "supervised self-regulation or 'co-regulation,' to direct regulation by an entity such as the FTC."<sup>60</sup> But based on our experience with the development of the Do Not Call registry and the practical problems that consumers face, it is EPIC's view that for a browser-based Do Not Track system to be successful, a centralized enforcement mechanism would be required.

---

<sup>53</sup> Ryan Singel, *supra* note 46.

<sup>54</sup> *Id.*; see also "Do Not Track Explained," *supra* note 45.

<sup>55</sup> "Do Not Track Explained," *supra* note 45.

<sup>56</sup> *Id.*

<sup>57</sup> "Do Not Track: Universal Web Tracking Opt-Out," project run by researchers at the Stanford Law School Center for Internet and Society and the Security laboratory at the Stanford Department of Computer Science, [www.donottrackus.org](http://www.donottrackus.org)

<sup>58</sup> *Id.*

<sup>59</sup> Cecilia Kang, "What a Do Not Track Option Might Look Like," The Washington Post Tech Blog, Nov. 17, 2010.

<sup>60</sup> "Do Not Track Explained," *supra* note 45.

The FTC recently released a privacy report that endorsed a Do Not Track mechanism but stopped short of discussing how such an approach would be made effective.<sup>61</sup> The report asks for comments on how Do Not Track would be implemented, but does explain that the most "practical method . . . would likely involve placing a setting similar to a persistent cookie on a consumer's browser and conveying that setting to sites that the browser visits." The FTC report also states that "there must be an enforceable requirement that sites honor those choices" but is vague on the details of how such enforcement would occur.

In EPIC's view, the FTC discussion of the Do Not Track proposal should have paid much closer attention to the history of Do Not Call. The agency has, in effect, attempted to replicate a successful program, Do Not Call, without recognizing the steps that were required to make the program work.

#### **IV. Issues with Do Not Track that Must Be Addressed**

##### *A. Opt-Out vs. Opt-In*

Individuals' rights and privacy would be more effectively protected by an opt-in framework rather than the opt-out do not track list being considered. An opt-in approach would require online advertisers and tracking companies to obtain express consent before tracking individuals.

An opt-in framework would better protect individuals' rights and is consistent with most United States privacy laws. For instance, the Family Educational Rights and Privacy Act, Cable Communications Policy Act, Electronic Communications Privacy Act, Video Privacy Protection Act, Driver's Privacy Protection Act, and Children's Online Privacy Protection Act all empower the individual by specifying that affirmative consent is needed before information is employed for secondary purposes.<sup>62</sup>

Opt-in is more effective than opt-out because it encourages companies to explain the benefits of information sharing, and to eliminate barriers to exercising choice. Experience with opt-out has shown that companies tend to obfuscate the process of exercising choice, or that exemptions are created to make opt-out impossible. For instance, the Gramm-Leach-Bliley Act required opt-out notices to be sent to customers of banks, brokerage houses, and insurance companies.<sup>63</sup> These notices were confusing and incomprehensible to many Americans.<sup>64</sup> Opting-out often required the consumer to send

---

<sup>61</sup> "Protecting Consumer Privacy in an Era of Rapid Change: A Proposed Framework for Businesses and Policymakers," Preliminary Federal Trade Commission Staff Report, p.66, December 2010, *available at* <http://ftc.gov/os/2010/12/101201privacyreport.pdf>.

<sup>62</sup> Respectively, at 20 U.S.C. § 1232g(b)(2)(A), 47 U.S.C. § 551(c)(2), 18 U.S.C. § 2511(2)(c), 18 U.S.C. § 2710(b)(2)(B), 18 U.S.C. § 2721(b)(11), and 15 U.S.C. § 6502(b)(A)(ii).

<sup>63</sup> 15 U.S.C. § 6801.

<sup>64</sup> Mark Hochhauser, "Lost in the Fine Print: Readability of Financial Privacy Notices," Privacy Rights Clearinghouse, July 2001, *available at* <http://www.privacyrights.org/ar/GLB-reading.htm>.

a separate letter to the company. Even if a consumer did opt out under the law, a company that wished to share consumer data could simply create a joint marketing agreement with another company to fall within an exemption to the prohibition on information sharing.<sup>65</sup>

In other contexts, phone companies have thwarted opt-out processes by demanding excessive authentication for opting out. For instance, the opt-out process for Customer Proprietary Network Information (CPNI) data sharing established by Verizon was confusing, and placed the burden on individuals to navigate a five-step process in order to opt-out.<sup>66</sup> Often, notices to consumers are not clear and therefore consumers are not making a meaningful choice when deciding whether to opt-out.<sup>67</sup>

While it seems that Do Not Track may end up being largely an opt-out type of mechanism, the idea that at least some data should be subject to consumers having to opt-in to have it collected should be considered, especially for sensitive health and financial information. If opt-out is the preferred strategy for Do Not Track, then it will require all of the elements that were eventually brought together for Do Not Call – centralized administration, enforceable legal protections, and a simple, transparent, and stable method for consumers to express their opt out preferences.

### *B. Opt-Out Cookies*

It is also important that Do Not Track is not based on the idea of opt-out cookies, such as those advocated by the NAI.<sup>68</sup> Opt-out cookies have been used before as mechanism for consumers to opt-out of being tracked, but they have not generally been successful. Opt-out cookies are a confusing and misleading approach to consumer privacy. They are counter-intuitive, as users concerned with privacy typically delete cookies, especially those associated with search activities.<sup>69</sup> Yet once the cookie is deleted, the privacy setting is lost and advertisers will no longer honor the user's privacy status.<sup>70</sup> Second, the opt-out cookie does not scale. If users are required to accept opt-out cookies for every site that they do not want tracking them, a person would have to keep cookies for every single Internet site, which does not make sense.<sup>71</sup>

---

<sup>65</sup> 15 U.S.C. § 6802 (b)(2).

<sup>66</sup> See Letter from Marc Rotenberg, Executive Director, Electronic Privacy Information Center, to Ivan Seidenberg, President and co-CEO, Verizon (Feb. 7, 2002), *available at* <http://www.epic.org/privacy/cpni/verizonletter.html>.

<sup>67</sup> See, e.g., FTC, "Transcript of December 7, 2009, Privacy Roundtable," Remarks of Alessandro Acquisti, Associate Professor, Carnegie Mellon University, Heinz College, *available at* [http://www.ftc.gov/bcp/workshops/privacyroundtables/PrivacyRoundtable\\_Dec2009\\_Transcript.pdf](http://www.ftc.gov/bcp/workshops/privacyroundtables/PrivacyRoundtable_Dec2009_Transcript.pdf) ("However, I see notification, control, and transparency as necessary conditions, but insufficient. . . . There is by now a wealth of behavioral data and databases showing what are the gaps between what consumers want in terms of privacy and their ability to achieve these stated intentions.").

<sup>68</sup> See, *infra* p. 6.

<sup>69</sup> Letter from EPIC et al. to Jim Lanzone, CEO Ask.Com, Dec. 20, 2007.

<sup>70</sup> *Id.*

<sup>71</sup> *Id.*

The browser-header approach to Do Not Track seems to eliminate this concern, as it is not cookie-based, but rather browser-based.

### *C. What Information is Collected?*

In any form of Do Not Track that it is implemented it is important to ensure that users are not required to give up private information in order to be on the registry or to use the browser-based mechanism. If an e-mail or IP address is collected, that could pose privacy concerns. Congress should investigate further what information the browser is sending back to companies in the "header" telling them that a user does not want to be tracked.

For example, the Ask Eraser product, which used opt-out cookies, inserted the exact time that a user enabled its product into the information that it sent in the browser.<sup>72</sup> The text string then operates like a unique identifier, such as a person's cellphone number or a social security number. While it is conceivable that there could be more than one cookie issued at the exact same second, it seems unlikely. Particularly, when histories are logged, reconstructing actual identity would be trivial. Also, even if Ask were not logging search histories, by transferring this type of cookie to third parties, it becomes easy for third parties to track users who have enabled Ask Eraser by simply noting the date/time stamp assigned.<sup>73</sup>

Therefore, any Do Not Track mechanism should be very cautious about what content is actually sent in the browser header to the online advertisers, and should ensure that it does not contain any information that can identify a user.

### *D. Tiered Web and Discrimination*

The worst form of privacy discrimination is to make access to information conditional upon the relinquishment of personal information. There is a possibility that Do Not Track could lead to a tiered web, that is, one where those who use Do Not Track can only see certain content. Whether this will happen depends on how online advertisers react to Do Not Track, but there is some evidence to suggest that a tiered web will not necessarily result.

Currently, users can implement ad blocking through a browser plug-in, and many do, but very few sites refuse to provide content to users who have enabled ad blocking.<sup>74</sup> And ad blocking would be much more costly to advertisers as it prohibits all ads, as opposed to Do Not Track, which would only prevent behavioral ads.<sup>75</sup> Additionally, a tiered web already exists in the form of those who are logged in when they browse versus those who are anonymous. It is unlikely though that disabling Do Not Track as a

---

<sup>72</sup> *Id.*

<sup>73</sup> *Id.*

<sup>74</sup> "Do Not Track Explained," *supra* note 45

<sup>75</sup> *Id.*

requirement for service or access to content will ever become as popular as requiring login.<sup>76</sup>

Obviously this would be a major concern if those using Do Not Track are blocked from accessing web content. Part of the enforcement mechanism surrounding Do Not Track should include penalties for any websites that engage in this kind of discrimination.

### *E. Preemption*

Congress should ensure that any Do Not Track legislation does not preempt state laws in the area of regulation of online data collection and targeted advertising. States have a traditional role in regulating privacy that should be preserved. There is a presumption in American law that state and local governments are primarily responsible for matters of health and safety.<sup>77</sup> Privacy is included in the category of health and safety issues, as an area of regulation historically left to the states.<sup>78</sup>

Federal consumer protection and privacy laws, as a general matter, operate as regulatory baselines and do not prevent states from enacting and enforcing stronger state statutes. The Electronic Communications Privacy Act,<sup>79</sup> the Cable Communications Privacy Act,<sup>80</sup> the Video Privacy Protection Act,<sup>81</sup> the Employee Polygraph Protection Act,<sup>82</sup> the Driver's Privacy Protection Act,<sup>83</sup> the Health Insurance Portability and Accountability Act,<sup>84</sup> the Gramm-Leach-Bliley Act,<sup>85</sup> and portions of the Fair Credit Reporting Act<sup>86</sup> all allow states to craft protections that exceed federal law. In each of the areas regulated by the above-referenced privacy laws, business has continued to flourish in states that have enacted privacy protections that are stronger than the federal law.

Permitting states to regulate interstate telemarketing will continue to promote regulatory innovation and experimentation. States enjoy a unique perspective that allows them to craft innovative programs to protect consumers. State legislators are closer to their constituents and the entities they regulate. Federal preemption can dilute more

---

<sup>76</sup> *Id.*

<sup>77</sup> *Hillsborough County v Automated Med. Labs.*, 471 U.S. 707, 716 (1985) (there is a "presumption that state and local regulation of health and safety matters can constitutionally coexist with federal regulation).

<sup>78</sup> *See, e.g., Hill v. Colo.*, 530 U.S. 703 (2000) (upholding a law protecting the privacy and autonomy of individuals seeking medical care, as the law was intended to serve the "traditional exercise of the States' police power to protect the health and safety of their citizens." (internal quotation marks omitted).

<sup>79</sup> 18 U.S.C. § 2710(f)(2005)

<sup>80</sup> 47 U.S.C. § 551(g) (2005)

<sup>81</sup> 18 U.S.C. § 2710(f) (2005).

<sup>82</sup> 29 U.S.C. § 2009 (2005).

<sup>83</sup> 18 U.S.C. § 2721 (2005)

<sup>84</sup> 29 U.S.C. § 191 (2005)

<sup>85</sup> 15 U.S.C. § 6701 (2005)

<sup>86</sup> 15 U.S.C. § 1681t (2005).

vigorous protections and policy debates that occur at the state level. For example, in a detailed study of caller ID policy approaches, researchers found that the FCC's position was much weaker than those developed by the states.<sup>87</sup> State and local governments are also more accountable than the federal government to their constituents. As a result, it is likely that stronger protections will emerge and more vigorous enforcement will be pursued by state actors.

Businesses are not put at a disadvantage by having to comply with differing state laws. In fact, businesses have long accommodated themselves to a range of state consumer protection statutes while maintaining a profitable enterprise. Courts have, for years, engaged in a process of reconciling potentially or actually conflicting laws through application of established legal principles to various factual situations. Such a tailored response is especially appropriate with respect to evolving technologies and new applications of those technologies. This flexible approach accommodates the needs of both businesses and consumers, while preserving state sovereignty in an area where states have traditionally had a significant role.<sup>88</sup>

#### *F. Enforcement*

As discussed earlier, this Do Not Track mechanism would need to be enforced by an agency such as the FTC.<sup>89</sup> And the enforcement must have teeth, otherwise it will not be at all effective. In addition to meaningful oversight by a federal agency, there should also be a private right of action that gives individuals, whose rights have been violated, the opportunity to seek relief. A private right of action is necessary even where a federal agency is given enforcement authority. Agency action is always discretionary and there is no guarantee, absent a private right of action, that an individual whose rights may have been violated will have the opportunity for relief. This problem has become even more evident in the least few years with the spotty record of the current FTC on matters concerning the protection of consumer privacy.

### **V. Conclusion**

Online data collected and targeted behavioral advertising pose a serious threat to consumer privacy. A Do Not Track mechanism, while important, only starts to solve one of the many problems with online data collection. EPIC respectfully requests the Committee to fully consider all of the issues with Do Not Track outlined in this statement, as well as the relevant history of the TCPA and Do Not Call list. A Do Not Track list can be an important tool, but only if it is done thoughtfully and enforced fully. At a minimum, EPIC believes that key to an effective Do Not Track technique will be the adoption of legislation that makes the decision by consumers enforceable, stable, transparent, and simple.

---

<sup>87</sup> Comments of EPIC et al to FCC regarding "Rules and Regulations Implementing the Telephone Consumer Protection Act of 1991," July 29, 2005, at 9.

<sup>88</sup> See, e.g., The national Association of Attorneys General Privacy Subcommittee, "Privacy Principles and Background," available at <http://www.naag.org/naag/resolutions/subreport.php>.

<sup>89</sup> See *supra* Part III.C

We would also strongly urge the Committee to undertake a more thorough examination of the Commission's strategy for safeguarding consumer privacy. In many areas, we believe the FTC has failed to take necessary steps to address clear public concerns about the collection and use of personal data for commercial purposes.

Thank you for your consideration of these views.

The Internet Society works to ensure the continued existence of a healthy Internet ecosystem. This includes support for multi-stakeholder activities that are open, inclusive, and generative. Key to this effort is the need to understand the complex balance between issues such as privacy, security, and reliability. When balanced properly, the result is a trusted network in which all participants, including users, enterprise and governments, have confidence using.

The organic growth of the Web as an effective means of communication over the Internet has given rise to uses of the technology that go beyond what was initially intended. Each innovation has provided the opportunity for both positive and negative, often unintended, consequences. One such set of trade-offs can be found in the pervasive use of mechanisms deployed to track users across the Web.

When a user directs a web browser to a specific site to request a page of content, there is a general (though often vague) understanding that the data between the end points moves through an unknown number of intermediaries (e.g. routers). Users, however, often operate with an implicit expectation that the persistent details of their interaction are limited to the two end points (i.e. the user and the known server). This is in strict contrast to the current norm in web browsing: each site often logs page content being retrieved, and a page is often a composite of content served from a number of additional end points (a.k.a. “third parties”). Each of these end points, in turn, is able to track various details regarding the user (e.g. their browsing, IP-based geo-location, etc.).

Increasing reliance on the Internet and related tools, such as the Web, is catalyzing demand for harmonized and interoperable privacy and data protection. A key component of the approach is the development of international legal frameworks. As part of this effort, policymakers are looking to technology, industry codes of conduct, certification schemes, and user education to compliment the emerging frameworks.<sup>i</sup>

Web tracking is receiving particular attention. For example, the Preliminary Federal Trade Commission Staff Report (December 2010) entitled *Protecting Consumer Privacy in an Era of Rapid Change – A Proposed Framework for Businesses and Policymakers*<sup>1</sup> states, among other things:

... Commission staff supports a more uniform and comprehensive consumer choice mechanism for online behavioral advertising, sometimes referred to as “Do Not Track.” Such a universal mechanism could be accomplished by legislation or potentially through robust, enforceable self-regulation. The most practical method of providing uniform choice for online behavioral advertising would likely involve placing a setting similar to a persistent cookie on a consumer’s browser and conveying that setting to sites that the browser visits, to signal whether or not the consumer wants to be tracked or receive targeted advertisements. To be effective, there must be an enforceable requirement that sites honor those choices.

Such a mechanism would ensure that consumers would not have to exercise choices on a company-by-company or industry-by-industry basis, and that such choices would be persistent. It should also address some of the concerns with the existing browser mechanisms, by being more clear, easy-to-locate, and effective, and by conveying directly to websites the user’s choice to opt out of tracking. ...

Supporting these efforts, research shows that users frequently respond to survey questions stating they do not want their browsing data to be collected without their knowledge and consent. A common conclusion from many surveys (including ones from The Annenberg Public Policy Center at the University of Pennsylvania, The Samuelson Law, Technology & Public Policy Clinic at UC Berkeley, and The PEW Internet & American Life Project) is that users want more transparency about data being collected, its use, and to have more control over it.

---

<sup>1</sup> <http://www.ftc.gov/os/2010/12/101201privacyreport.pdf>



Missing from the analysis of these surveys, however, is consideration of how users expect to effectively balance all of the related issues around increased privacy controls. It is unclear how users will react when privacy is increased with a related impact on security, usability, and reliability. Historically, when considering adoption of security technologies, average users opt for the simplest experience, even when it is the least secure.

To fill out the picture further, there are various reasons to employ mechanisms for tracking Web users. There are also various methods by which users can be tracked. Some methods include cookies (browser-based or managed by add-ons such as Adobe Flash), others rely on browser fingerprinting (i.e. using unique characteristics in response headers), while still others leverage network and device characteristics (e.g. IP addresses and MAC identifiers).

Regardless of the reason for tracking users or the method used, tracking falls into one of two classes:

- **Single-Site Tracking** – There is a “first-party” relationship between the user and the known site. Activities are being tracked, sometimes unknowingly, but the resulting data is managed for the use of the site itself.
- **Multi-Site Tracking** – In contrast to single-site tracking, users are tracked across sites and by multiple sites. This introduces one or more third parties to the interaction between the user and the known site.

A common use of tracking for a single site is to observe and monitor the interactions of users within their service. A goal is to compare similar users in an effort to personalize the user experience on the site (a.k.a. “behavioral profiling”). Another related use is to improve the effectiveness of display advertising by observing and analyzing user patterns across multiple sites (a.k.a. “behavioral advertising”). In addition to content and service delivery, another common use of tracking is to improve security by monitoring user activities (e.g. building behavioral risk profiles).

Some consideration should also be given to differences between tracking methods used within the context of browsing activities and those used for the business of brokering user data. In one case, regardless of how the tracked data is collected (on a single site or across multiple sites), it is analyzed and used only by the collector and its agents. In other cases, the collector may share with or sell to other (often undisclosed) parties (a.k.a. second parties) the data that is collected. It is important when considering issues around tracking users to be aware of both modes, understanding that they may also work in conjunction.

Given the rapid expansion of the Web into all aspects of daily life, it is clear that issues of online privacy need to be addressed, while not adversely affecting the overall utility of the Internet. Protecting user privacy online cannot be taken lightly, and requires well-considered solutions that are open, transparent, and inclusive.

*This paper was prepared by Christine Runnegar (runnegar@isoc.org) and J. Trent Adams (adams@isoc.org) for the purpose of participating in the W3C “Workshop on Web Tracking and User Privacy” at the Center for Information Technology Policy at Princeton University in Princeton, NJ, USA (28-29 April 2011)*

i. Some examples of recent international and regional privacy initiatives:

The OECD “is preparing an anniversary report on the evolving privacy landscape” (see [http://www.oecd.org/document/35/0,3746,en\\_2649\\_34255\\_44488739\\_1\\_1\\_1\\_1,00.html](http://www.oecd.org/document/35/0,3746,en_2649_34255_44488739_1_1_1_1,00.html))

In Europe, the Council of Europe is considering how to modernize the *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data* (Convention 108) (see [http://www.coe.int/t/dghl/standardsetting/DataProtection/default\\_en.asp](http://www.coe.int/t/dghl/standardsetting/DataProtection/default_en.asp)) and the European Commission is in “... the process of reviewing the general EU legal framework on the protection of personal data” including *Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data* (see [http://ec.europa.eu/justice/policies/privacy/review/index\\_en.htm](http://ec.europa.eu/justice/policies/privacy/review/index_en.htm)).

APEC economies, through the APEC Data Privacy Pathfinder, are “... develop[ing] and test[ing] the essential practical elements of a system that would enable accountable cross-border data flows under the guidance of APEC data privacy principles” (see <http://www.apec.org/en/Groups/Committee-on-Trade-and-Investment/Electronic-Commerce-Steering-Group.aspx>)

In 2009, the 31<sup>st</sup> International Conference of Data Protection and Privacy Commissioners produced a *Joint Proposal for a Draft of International Standards on the Protection of Privacy with regard to the processing of Personal Data* (“the Madrid Resolution”) (see International Standards on the Protection of Personal Data and Privacy at <http://www.justice.gov.il/PrivacyGenerations/adopted.htm>).

In 2010, the 32<sup>nd</sup> International Conference of Data Protection and Privacy Commissioners adopted a *Resolution calling for the organisation of an intergovernmental conference with a view to developing a binding international instrument on privacy and the protection of personal data* (see Resolution on International Conference at <http://www.justice.gov.il/PrivacyGenerations/adopted.htm>).

# Browser Vendors: fight for your users

Thomas Lowenthal  
Center for Information Technology Policy  
Princeton University  
tlowenth@princeton.edu

## Abstract

The parties who track users online are technically sophisticated, dedicated, and motivated by significant financial gains. Users often lack the technical knowledge to understand the forms of tracking that are deployed against them, the skills necessary to deploy countermeasures, or the significant quantities of time and effort necessary to safeguard their privacy. Browser vendors, on the other hand, have the resources, capacity, and expertise necessary to protect their users from many different privacy threats. Browser vendors should take responsibility for their role as users' agents online and use their technical and market power to protect user interests.

## 1 The Browser: Users' Privacy Trust Root

The web browser is literally the user's representative online. As the user's agent, a browser should act based on the needs of the user; design policy decisions should be based exclusively on the user's priorities. Indeed, the browser is the only party that the user should have to rely upon to work for them: it's much easier to make a one-time trust judgment about which web browser to use than it is to have to make repeated, ongoing, granular trust judgments about numerous websites, and their embedded and active content.

Indeed, it would be prohibitive to expect users to audit the potential privacy risks posed by the embedded web bugs, persistent & novel cookies, JavaScript content, tracking practices, and information sharing

policies of all the many sites they visit. The much more reasonable model has the user choose a trustworthy browser, learn about its security and privacy features, customize individual settings, and then confidently rely that the browser will work to protect them in the choices that they've made, and will make ongoing operational decisions based on the user's expressed preferences.

Browser vendors should protect their users by making privacy-by-design a priority the same way that they do with security. In addition, browsers should be honest with their users, explaining their strengths and weakness, so that users can make informed activities about their activities online. What follows is a selection of ways that browsers currently fail to protect their users' privacy. The difficulty of mitigating or fixing these problems varies, but browser vendors should consider these issues — and others like them — to be important ways that they can protect (or fail to protect) their users. Because of the browser's unique position in users' web-browsing trust hierarchies, these issues demand fixes at the browser level.

## 2 Web Privacy Weaknesses & Countermeasures

### 2.1 Cookie & Active Tracking Control

Most users are aware of HTTP cookies, and some are aware of other active tracking measures like flash cookies. However, there are many[7] active tracking measures that can be used to identify and re-identify users. Many of these were not even designed as identification technologies, but result from the 'generous' set of features available among the variety of browser and active content technologies available on the web.

Given how much of our lives we spend online, persistent and pervasive tracking poses a direct threat to individual privacy. It's not that tracking technologies are inherently wrong, far from it. Rather browsers should offer users the technical capacity to choose which sites know and retain what information about them, over which sessions. Defaulting to letting sites keep persistent, hard-to-remove track of users is a mistake: tracking should be an option that's up to the user, and under their control

## 2.2 Fingerprint Uniqueness Reduction

Even when not using active tracking methods like cookies, passive tracking methods often allow for accurate re-identification of a particular browser. According to the data produced by the EFF's PanoptiClick project[3], browsers' fingerprints have an average anonymity set size larger than 280,000, and browsers supporting Flash or Java are 94.2% likely to be unique. However, in the custom browser deployed by the Tor Project, the measures taken to create a uniform browser fingerprint were quite successful, producing highly uniform anonymity sets.

There are lots of trivial steps that browser vendors can take to protect against this method for identifying users. Reporting a slightly more granular browser version number like "1.6" rather than "1.6.0.17" immediately makes fingerprints more homogeneous. Likewise, sorting supported font lists before reporting them takes away another significant source of entropy. These are just some changes made based on the entropy data. Browser vendors have the ability to reconsider the amount of information they really *need* to report to sites. Defaulting to reporting everything may be somewhat sensible in a fragmented, browser-dependent web. However, in a web built on agreed standards, privacy should be the default, with exceptions made for specific information when needed.

## 2.3 Effective Private Browsing Modes

Most of the modern browsers feature private browsing modes, but research from Stanford University[1] suggests that they may not be well-implemented to provide the sort of privacy protections that users might expect. In addition to exploitable weaknesses which may allow traces to be left locally after private browsing, these modes fail to implement the

anonymity measures which would be required to prevent a hostile website from associating non-private browsing with a series of distinct private browsing sessions.

Private browsing modes are an important tool in a users's privacy defense arsenal. They allow users to retain control of their personal information in ways which might not otherwise be possible. They may even permit users to engage in behavior which they might otherwise have considered too risky. As such, it's imperative that these modes are effective, and live up to users' functionality expectations.

## 2.4 History Retrieval

It has for some time been possible to use cunningly crafted HTML & CSS to infer users' complete browsing history[6], which may contain all kinds of sensitive information, and — moreover — makes for a fairly unique way to re-identify the same users. This is mentioned less to draw attention to this particular attack, and more as a comment on these sorts of browser weaknesses. As long as browser vendors leave this sort of gaping vulnerability unchecked, their users will continue to be at risk.

The problem is that the drive to patch privacy holes doesn't seem to be nearly as strong as the drive to fix security holes, or developing new and innovative features. However, for many users, improved privacy protection is much more valuable than shiny new tab-sorting features. While competitions like Pwn2Own glamourize and reward security development, privacy design often plays second fiddle.

## 2.5 Certificate Trust Control

As recent events[2][8][5] and commentary[10][4][9] have indicated, the public-key identification infrastructure which underpins our web encryption technology is hopelessly broken. This failure isn't a technical one, it's a social one, and browser vendors are at least partly to blame. There have been no movements to revoke the signing powers of the several certificate authorities which fail. Users rely on the the security practices of every single certificate authority whenever they do online banking, or transfer personal medical information online. When a CA spectacularly fails, a browser vendor should pro-actively call them on it, acting on the trust that users place in their browser by revoking the CA's authority.

Yes, these sorts of aggressive enforcement actions ‘break’ some sites. However, that should be the desired behavior. When the browser represents to the user that a secure connection is taking place, it should be on the basis of that actually being true. If a CA is failing their authentication responsibility, the browser should not mislead the user by asserting that everything is hunky-dory when an attack may actually be taking place.

### 3 Conclusion

The browser is the user’s only intermediary and protector from the dangerous ravages of a cold, dark, unfriendly web. It is practically the case that web services lust after users’ personal information, extended click- and browsing-history, and mostly succeed in getting it. A browser sits as the root of a user’s trust tree, and has a unique responsibility to safeguard the user’s privacy interests online.

External policy measures like Do Not Track, data breach notifications, privacy policies, and personal information protection laws are valuable, but they have their limitations. Laws are hard to enforce across borders; privacy policies are incredibly difficult to read and even harder for users to verify or audit. The best way to keep information from being used against the user is to prevent it from leaking out in the first place. That begins and ends with the browser.

### References

- [1] Gaurav Aggarwal et al. “An Analysis of Private Browsing Modes in Modern Browsers”. In: *USENIX 2010*. 2010. URL: <http://crypto.stanford.edu/~dabo/pubs/papers/privatebrowsing.pdf>.
- [2] Jacob Appelbaum. “Detecting Certificate Authority compromises and web browser collusion”. In: *The Tor Blog* (Mar. 22, 2011).
- [3] Peter Eckersley. “How Unique Is Your Web Browser?”. In: *Privacy Enhancing Technologies Symposium (PETS 2010)*. Ed. by Electronic Frontier Foundation. 2010. URL: <https://panopticlick.eff.org/browser-uniqueness.pdf>.
- [4] Ed Felten. “Web Certification Fail: Bad Assumptions Lead to Bad Technology”. In: *Freedom to Tinker* (Feb. 23, 2010). URL: <http://www.freedom-to-tinker.com/blog/felten/web-certification-fail-bad-assumptions-lead-bad-technology>.
- [5] Phillip Hallam-Baker. “The Recent RA Compromise”. In: (Mar. 23, 2011). URL: <http://blogs.comodo.com/it-security/data-security/the-recent-ra-compromise/>.
- [6] Artur Janc and Lukasz Olejnik. “Feasibility and Real-World Implications of Web Browser History Detection”. In: *Web 2.0 Security and Privacy 2010*. 2010. URL: <http://w2spconf.com/2010/papers/p26.pdf>.
- [7] Samy Kamkar. *Evercookie*. URL: <http://samy.pl/evercookie/>.
- [8] Steve Schultze. “Web Browsers and Comodo Disclose A Successful Certificate Authority Attack, Perhaps From Iran”. In: *Freedom to Tinker* (Mar. 23, 2011). URL: <http://www.freedom-to-tinker.com/blog/sjs/web-browsers-and-comodo-disclose-successful-certificate-authority-attack-perhaps-iran>.
- [9] Steve Schultze. “Web Security Trust Models”. In: *Freedom to Tinker* (Feb. 22, 2010). URL: <http://www.freedom-to-tinker.com/blog/sjs/web-security-trust-models>.
- [10] Christopher Soghoian and Sid Stamm. “Certified Lies: Detecting and Defeating Government Interception Attacks Against SSL”. In: *SSRN* (Apr. 14, 2010). URL: [http://papers.ssrn.com/sol3/papers.cfm?abstract\\_id=1591033](http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1591033).

## Do Not Track as a Generative Approach to Web Privacy

Jonathan Mayer<sup>1</sup>

Consider behavioral advertising as a hypothetical negotiation problem.<sup>2</sup> On one side of the table is the average user, who wants to access an advertising-supported service—but only give up some privacy in exchange.<sup>3</sup> On the other side is the average online business, glad to provide a service to the user—if able to display an ad, and preferably an interest-targeted one.<sup>4</sup> In the status quo the user is tracked, and the site delivers an interest-targeted ad: the user gets her least preference, and the site gets its greatest preference.<sup>5</sup> But suppose the site could deliver a privacy-preserving interest-targeted ad. The user would be better off, and the site would be no worse off.<sup>6</sup>

Technologies exist for privacy-preserving interest-targeted advertising—they just haven't been adopted.<sup>7</sup> This paper argues that privacy-friendly advertising and similar gains could be achieved by moving privacy choices to a generative platform, and it shows how Do Not Track will do just that.

---

<sup>1</sup> Ph.D. & J.D. student, Stanford University; Student Fellow, Stanford Center for Internet and Society.

<sup>2</sup> This discussion is greatly simplified for clarity. Some users are accepting of third-party tracking. The hypothetical omits the role of advertising networks, defines the status quo as solely behavioral advertising, and assumes that a site marginally prefers to display a behavioral ad. For an empirical analysis of these issues, see Jonathan Mayer, *Do Not Track Is No Threat to Ad-Supported Businesses*, CENT. FOR INTERNET & SOCIETY (Jan. 20, 2011), <http://cyberlaw.stanford.edu/node/6592>.

<sup>3</sup> Studies have consistently shown that users overwhelmingly reject third-party web tracking. See, e.g., E.g., Joseph Turow et al., *Americans Reject Tailored Advertising and Three Activities that Enable It* 15 (Sept. 29, 2009), available at <http://ssrn.com/abstract=1478214>; Lymari Morales, *U.S. Internet Users Ready to Limit Online Tracking for Ads*, GALLUP (Dec. 21, 2010), <http://www.gallup.com/poll/145337/internet-users-ready-limit-online-tracking-ads.aspx>.

<sup>4</sup> See Mayer, *supra* note 2.

<sup>5</sup> See Julia Angwin, *The Web's New Goldmine: Your Secrets*, WALL ST. J., July 30, 2010.

<sup>6</sup> All else being equal, of course.

<sup>7</sup> E.g., Vincent Toubiana et al., *Adnostic: Privacy Preserving Targeted Advertising*, PROC. 17TH ANN. NETWORK & DISTRIBUTED SYS. SECURITY SYMP. (2010), available at <http://crypto.stanford.edu/adnostic/adnostic-ndss.pdf>; Matthew Fredrikson & Ben Livshits, *RePriv: Re-Envisioning In-Browser Privacy* (Microsoft Research Technical Report MSR-TR-2010-116, 2010), available at <http://research.microsoft.com/pubs/137038/tr.pdf>.

## **The Platform for Privacy Preferences (P3P)**

The notion of a privacy negotiation is nothing new.

The original web suffered from amnesia. Quit your browser and every interactive site was reset. And so, in 1994, a Netscape engineer implemented a fix: the cookie, a remotely accessible data store within the browser.<sup>8</sup>

Just three years later, every major browser supported cookies. Users could save shopping carts; they could store preferences; and they could maintain a login. But users' activities also could be—and increasingly were—tracked, not only by the sites they visited but also by invisible third parties.

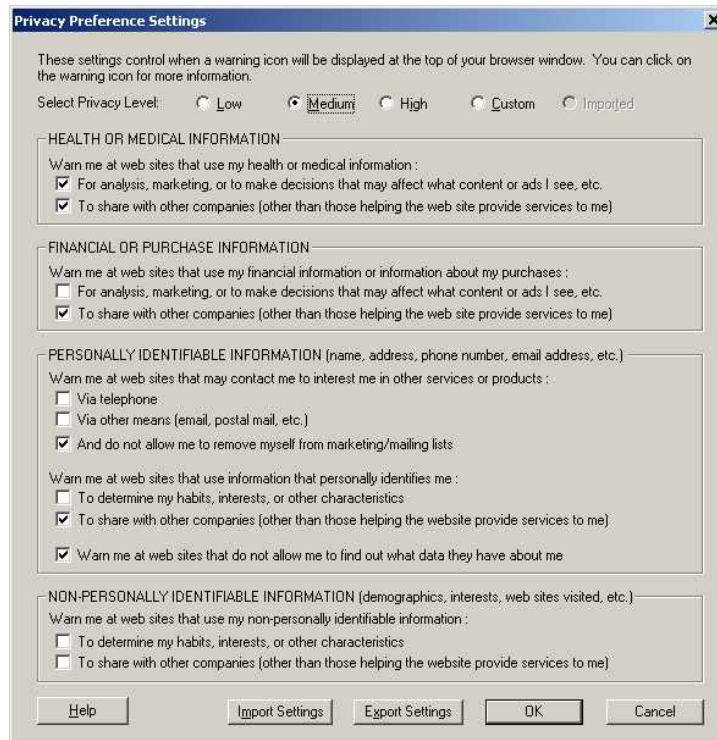
Recognizing the privacy threat, a group of concerned computer scientists began work on the Platform for Privacy Preferences (P3P), a technical mechanism for a privacy negotiation between a user and a website. A user would declare her privacy preferences to her browser, and a site would declare its privacy policy in a computer-interpretable form. Upon visiting a site, the browser would match the user's preferences to the site's policy. If the two aligned, the browser would load the site. If not, the user would have a choice of whether to allow the site anyways or use site-specific, issue-by-issue opt outs.

The protocol specification aimed to be sufficiently fine-grained and flexible to capture the nuance of privacy policies. A site could, for example, indicate it would share a user's ZIP code, pager number, and political affiliations with an advertising network, but keep to itself her age, employer, and health records. Likewise a user could fine-tune privacy preferences, such as allow sites to share purchase history and general interests, but not financial information.

The P3P project intended to release a standard in eighteen months.<sup>9</sup>

---

<sup>8</sup> John Schwartz, *Giving Web a Memory Cost Its Users Privacy*, N.Y. TIMES, Sept. 4, 2001.



### P3P Browser Preferences<sup>10</sup>

Click + for more detailed information

**+ AT&T Privacy Practices**

**Privacy Policy Check**

- AT&T's privacy policy *does not match your preferences*:

- Site may use financial information or information about your purchases for analysis or to make decisions that may affect what content or ads you see, etc.
- Unless you opt-out, site may use financial information or information about your purchases for marketing

**Privacy Policy Summary**

**+ Policy Statement 1 - General**  
AT&T uses your personally identifiable information for billing purposes, to provide services to you, and to inform you of services that may better meet your needs, but we do not disclose your personally identifiable information to third parties who want to market products to you, period.

**+ Policy Statement 2 - Clickstream**  
We want to make the content on our sites as relevant, interesting and timely as possible and to do that we use information about which pages you visit on our site. AT&T uses advertising companies to deliver ads on some AT&T Web sites. The advertising companies may also receive some anonymous information about ad viewing by Internet users on AT&T Web sites. This information cannot be associated with a name or email address without the customer's permission.

**- Access to your information**  
This site allows you to access your contact information and certain other information about you from its records

**+ How to reach this site**  
**+ How to resolve privacy-related disputes with this site**

**More Information**

[Read this site's full privacy policy](#)  
[Find out how to opt-out](#)

### P3P Policy Warning<sup>11</sup>

<sup>9</sup> Platform for Privacy Preferences Project, *Project Update* (July 10, 1997), <http://www.w3.org/P3P/100797Update.html>.

<sup>10</sup> Privacy Bird, *Privacy Bird Tour*, [http://www.privacybird.org/tour/1\\_3\\_beta/tour.html](http://www.privacybird.org/tour/1_3_beta/tour.html).



It took five years; P3P was finally standardized in 2002.<sup>12</sup> But few tools existed for creating policies, only a minority of sites adopted P3P, and web browsers implemented only bits and pieces of the standard. After a final effort to reinvigorate the project, in late 2006 the P3P standards group unraveled.<sup>13</sup> Few P3P policies remain, and most do not conform to the standard.<sup>14</sup>

## Generativity and Privacy Choice

In the wake of P3P's failure, critics have launched a number of assaults: it presented users with far too many and too complex choices;<sup>15</sup> it was difficult to enforce;<sup>16</sup> and its language was inadequate for capturing the nuance of privacy policies.<sup>17</sup> All fair points. But here's one more, which I view as the most fatal: P3P was not generative.

In *The Future of the Internet—And How to Stop It* Jonathan Zittrain endeavored to identify the properties of technologies that lead to explosive, unguided innovation. He argued for five factors, technologies that<sup>18</sup>

- Make difficult tasks easier;
- Are easily adapted to new purposes;
- Require little to no expertise or training;

---

<sup>11</sup> *Id.*

<sup>12</sup> LORRIE CRANOR ET AL., THE PLATFORM FOR PRIVACY PREFERENCES 1.0 (P3P1.0) SPECIFICATION (Apr. 16, 2002), available at <http://www.w3.org/TR/P3P/>.

<sup>13</sup> LORRIE CRANOR ET AL., THE PLATFORM FOR PRIVACY PREFERENCES 1.1 (P3P1.1) SPECIFICATION (Nov. 13, 2006), available at <http://www.w3.org/TR/P3P11/>.

<sup>14</sup> Pedro Giovanni Leon et al., *Token Attempt: The Misrepresentation of Website Privacy Policies Through the Misuse of P3P Compact Policy Tokens*, PROC. 9TH ANN. ACM WORKSHOP ON PRIVACY IN THE ELECTRONIC SOC'Y (2010).

<sup>15</sup> Ari Schwartz, *Looking Back at P3P: Lessons for the Future* (Nov. 2009), available at [https://www.cdt.org/files/pdfs/P3P\\_Retro\\_Final\\_0.pdf](https://www.cdt.org/files/pdfs/P3P_Retro_Final_0.pdf).

<sup>16</sup> Ruchika Agrawal, *Why is P3P Not a PET?* (2002), <http://www.w3.org/2002/p3p-ws/pp/epic.pdf>.

<sup>17</sup> Lorrie Faith Cranor, *Incentives for Adoption of Machine-Readable Privacy Notices* (Nov. 5, 2010), [http://www.iab.org/about/workshops/privacy/papers/lorrie\\_cranor.pdf](http://www.iab.org/about/workshops/privacy/papers/lorrie_cranor.pdf).

<sup>18</sup> JONATHAN ZITTRAIN, THE FUTURE OF THE INTERNET—AND HOW TO STOP IT 71-73 (2008). See also James Grimmelman & Paul Ohm, Book Review, *Dr. Generative or: How I Learned to Stop Worrying and Love the iPhone*, 69 MD. L. REV. 910 (2010).

- Are easy to learn about and acquire; and
- Facilitate transfer of changes.

Zittrain bundled these properties into a solitary adjective: “generative.”

For a privacy choice platform to succeed, it must be generative. New websites, web services, web business models, and web technologies are established daily. As a consequence, web privacy considerations are in constant flux. How would an ossified, purpose-built privacy choice mechanism respond to content-sharing sites? Social networking? Social plug-ins such as the Like button? Single sign-on like OpenID? Would web businesses have to retain privacy platform consultants? Would there have to be associations and conferences just for privacy platform experts?

Such would have been P3P’s fate, if it had lasted longer. P3P was difficult to implement for a browser or website, narrowly purposed, convoluted, under-documented, and difficult to generalize across sites. It wasn’t generative. And so it failed.

### **Allocative Technologies**

Perhaps a generative privacy choice platform could be developed. I have doubts. But here’s an alternative approach: Instead of constructing a new generative platform, why not build on an existing one? And, when a problem does not naturally fall to the generative platform, why not use simple mechanisms—for convenience, “allocative technologies”—to relocate the problem there?<sup>19</sup>

---

<sup>19</sup> This argument suggests a rough technological parallel to Guido Calabresi’s “cheapest cost avoider” thesis: allocate a difficult online problem to the most generative system available.

Language signaling is a common allocative technology. Browsers don't include sophisticated translation software. Instead, they signal a user's language preferences, and it's up to foreign sites to develop alternate-language versions using standard web technologies.

Mobile web browsing now relies extensively on allocative technology. Before the iPhone, most mobile device browsers would attempt (unsuccessfully) to adapt websites for easier viewing on a small screen. Recognizing the failure of this approach, Apple launched its mobile browser with an explicit reliance on allocative technology: Apple encouraged websites to build mobile-friendly versions of their sites using standard, generative web technologies. In response to a request from an iPhone, sites were to redirect to their mobile versions. This allocative approach is so successful that every major mobile browser since has adopted it.

### **Do Not Track as an Allocative Technology for Privacy Choice**

Do Not Track is an allocative technology for privacy choice: it relocates the third-party privacy negotiation from the browser, where it has languished since P3P, to the web. In response to a Do Not Track user's request, a web service is free to respond using the standard web technology toolset. It could just deliver its service and an ad without tracking. Or it could ask a user for her interests to deliver a privacy-preserving interest-based ad. Or it could ask for a small payment. It could even refuse to provide service until the user disables Do Not Track.

And there, at last, is the long-sought web privacy negotiation. Do Not Track gives users a veto of the status quo, and allows web services to respond with meaningful privacy choices built on a generative platform.

# Web Tracking Protection

## W3C Workshop on Web Tracking and User Privacy

28/29 April 2011, Princeton, NJ, USA

[Adrian Bateman](#), Internet Explorer Program Manager, Microsoft Corporation

At Microsoft, we believe that it is critical for the industry to build innovative solutions founded on the principles of transparency, control and security. We are very proud that Internet Explorer 9 was the [first major browser to respond](#) to the recent call from the Federal Trade Commission in the United States for a “Do Not Track” mechanism.

In February, we made a [Member Submission](#) to the W3C proposing a [Web Tracking Protection](#) standard. This specification is designed to help users have better control over their online information and has two parts:

- Filter lists, which can enforce user privacy preferences by preventing the user agent from making unwanted requests to specific third party web servers that could be used to track users as they visit first party web sites.
- A “Do Not Track” user preference, which is an HTTP header and a DOM property.

Together these technologies can be used to provide privacy protections for users by helping them to control which third party web sites their user agent communicates with and to allow their user agent to signal their intention with respect to tracking.

A filter list contains parts of third-party URIs that a browser may access automatically when referenced within a web page that a user deliberately visits. Rules in a filter list may change the way the user agent handles third-party content. By limiting the calls to these third party web sites the filter list limits the information other sites can collect about a user.

The “Do Not Track” user preference is maintained by the user agent and is exposed as both a HTTP header, which can be read by a web server, and as a DOM property, which can be accessed by client-side JavaScript. The DOM property is particularly important in environments where the site developer may not have access to the raw HTTP request headers. Both the header and property convey the same user preference.

Web sites that choose to respect the “Do Not Track” user preference will read this value and will not “track” the user when this setting is enabled. This depends on what the definition of “tracking” is and what it means not to track a user. This is a complex topic as “tracking” could cover a wide variety of activities including online behavioural advertising, analytics, etc. Microsoft has submitted a separate paper about defining the meaning of “track” and appropriate web site responses to the “Do Not Track” user preference.

The final version of Internet Explorer 9 [released on March 14](#) implements full support for the features described in our Web Tracking Protection submission. With IE9, anyone on the Web can create and publish a Tracking Protection filter list. These simple files are uploaded to a web site and

made available to others via a link. Users can create or subscribe to more than one list if they wish and, because the Web evolves over time, IE9 will automatically check for updates to the user's lists on a regular basis. We invite workshop participants and the wider community to review the details of the submission.

## **References**

Web Tracking Protection

W3C Member Submission 24 February 2011

<http://www.w3.org/Submission/2011/SUBM-web-tracking-protection-20110224/>

# Tracking to Consensus: Coordination of Policy and Technical Standardization in Web Privacy Efforts

**W3C Workshop on Web Tracking and User Privacy  
28/29 April 2011, Princeton, NJ, USA**

Sue Glueck, Senior Attorney, Microsoft Corporation and Craig Shank, General Manager, Interoperability Group, Microsoft Corporation

At Microsoft, consumer trust is vital to our business, and privacy is a critical component of earning and maintaining that trust. In all of our service offerings, we strive to be transparent about our privacy practices, offer meaningful privacy choices, and protect the security of the data we store.

The explosive growth of the Internet, cloud computing, the proliferation of computers and handheld mobile devices, and the expansion of e-commerce, e-government, e-health, and other web-based services have brought tremendous social and economic benefits. At the same time, however, technology has fundamentally redefined how, where, and by whom data is collected, used, and shared. The challenge that industry, government, academics, and advocates must address together is how to best protect consumers' privacy while enabling businesses to develop a wide range of innovative products and services.

The multiple contexts in which Microsoft engages with consumers give us a unique perspective on the privacy discussion. For example, as a website operator, an ad network, and a browser developer, we have a deep understanding of the roles that different participants in the digital ecosystem play in safeguarding consumer privacy. Also, based on our longstanding involvement in the privacy debate, we recognize that the combined efforts of industry and government are required to effectively balance the need to protect consumers' privacy interests and promote innovation.

When Justice Louis Brandeis famously defined privacy as "the right to be let alone" in 1890,<sup>†</sup> he could not have foreseen how technology would revolutionize our world. In the digital era, privacy is no longer about being "let alone." Privacy is about knowing what data is being collected and what is happening to it, having choices about how it is collected and used, and being confident that it is secure. These three principles—transparency, control, and security—underpin Microsoft's approach to privacy. We believe

---

<sup>†</sup> Samuel D. Warren & Louis D. Brandeis, *The Right to Privacy*, 4 HARV. L. REV. 193, 193 (1890). Accessed at [http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy\\_brand\\_warr2.html](http://groups.csail.mit.edu/mac/classes/6.805/articles/privacy/Privacy_brand_warr2.html)

that the principles of transparency, control, and security should inform technological, self-regulatory, legislative, and educational initiatives to safeguard consumer privacy.

In a separate paper from Adrian Bateman entitled “Web Tracking Protection” we have written about the browser features in IE9 and Microsoft’s [Member Submission](#) to the W3C. Here we would like to identify some of the key questions we believe need to be addressed across the stakeholder communities in order to make any of these approaches effective.

In light of our experience, we continue to advocate a multi-pronged approach that includes technology tools – such as the Tracking Protection and “Do Not Track” user preference described in our W3C submission – as well as industry self-regulation, legislation, and consumer education. We have written in more detail about each of these in connection with the recent Senate Commerce Committee hearings [here](#).

For the purpose of this workshop and discussions of the role of W3C and other organizations in Web privacy and tracking protection, one of our key focus areas is effective coordination of different elements of these approaches, different stakeholder views, and the alignment of technical standards with policy interests.

Over the past ten years, there have been a number of thoughtful papers on the connection between technical standards and policy interests.<sup>5</sup> Fundamental to that discussion is a recognition that the work on technical standards designed to implement policy or “values” will need to integrate views that reach beyond the discussions that may take place in a strictly technical standardization effort. We believe that it is important for the discussions at the Workshop to move the conversation toward consensus on how some of the underlying “values,” “social protocols” or “policy and business rules” can be identified and developed in tandem with the technical means to achieve them.

---

<sup>5</sup> We appreciate the pointer from Deirdre Mulligan at the UC Berkeley School of Information to several of these, including:

- Nick Doty, Dierdre K. Mulligan and Erik Wilde, *Privacy Issues of the W3C Geolocation API*, UC BERKELEY SCHOOL OF INFORMATION REPORT 2010-038 (February 2010). Accessed at <http://escholarship.org/uc/item/0rp834wf#page-2>
- Lorrie Faith Cranor and Joseph Reagle Jr., *Designing a Social Protocol: Lessons Learned from the Platform for Privacy Preferences*, PROCEEDINGS OF THE TELECOMMUNICATIONS POLICY RESEARCH CONFERENCE (September 27-29, 1997). Accessed at <http://www.w3.org/TR/NOTE-TPRC-970930/>
- John Morris and Alan Davidson, *Policy Impact Assessments: Considering the Public Interest in Internet Standards Development*, TPRC 2003 – THE 31ST RESEARCH CONFERENCE ON COMMUNICATION, INFORMATION AND INTERNET POLICY (August 2003). Accessed at <http://www.cdt.org/publications/pia.pdf>

Accordingly, we believe that some of the key questions that should be discussed in this Workshop include:

- What is the appropriate process – including policy and broad stakeholder input – to develop the definition of “track” and web site behaviors in response to the “Do Not Track” signal from a browser?
- Who are the appropriate stakeholders to be engaged in developing that definition and behaviors?
- What will be the most effective way to convene those stakeholders in that development?
- What objectives, considerations, constraints and other factors should stakeholders have in mind as they look at potential approaches to web privacy – for example to determine what actions web sites should take in response to a “Do Not Track” signal?
- How will that process best take into account the global nature of the web – for example if a consumer in Brazil accesses a French web site running on servers hosted in Germany using an ad provider from Australia and an analytics firm in the United States, how do all of the participants in the system know what definition of tracking applies and how to interpret the consumer’s expression of intent?

We believe that a robust discussion of these questions will help move the overall efforts on the technology and the related policy topics forward. We also believe that a broad set of stakeholders is required to achieve effective outcomes on these issues for industry, government, and most importantly, consumers.





**Position Paper: Do Not Track  
W3C Workshop on Web Tracking and User Privacy, April 28-29, 2011  
Prepared by Mozilla and Submitted on March 25, 2011**

Mozilla supports a full range of innovations and industry practices that enhance consumer choice and control with regard to online behavioral advertising. This includes the creation of a uniform and comprehensive choice mechanism through a new Do Not Track (DNT) HTTP header as another step in a series of many privacy improvements. Continued leadership is required to develop consensus on the scope of DNT as it relates to online behavioral advertising and implementation across the online advertising industry. We are interested in participating in the upcoming W3C workshop to share our recent experience in implementing the DNT header in Firefox 4, how industry continues to rise to the occasion in crafting a response, as well as how we think the W3C efforts fit with our parallel submission to the IETF.

**Do Not Track Mechanisms for Online Behavioral Advertising**

Unlike blocking lists or opt-out cookies, which place the burden on the consumer and, more importantly, do not respond to all forms of OBA-related tracking and targeting, a DNT header has the potential for consumers to broadcast preferences for advertisers and publishers to honor while not undermining or blocking more widely-accepted and privacy-preserving forms of advertising. Success of the header approach will require support and collaboration from stakeholders across the web technology and display ad ecosystem.

Since the release of the FTC's proposed framework, there has been considerable public and media attention given to the topic of online behavioral advertising (OBA) and the FTC's recommendation for the creation of a Do Not Track (DNT) mechanism. Mozilla recently added the new HTTP DNT header that Firefox users can use to state a preference to not be tracked across websites for advertising. This feature easily co-exists with other browser-based privacy and cookie-based tools already available to Firefox users today.<sup>1,2,3</sup>

The DNT header builds on the work of the advertising networks by re-framing the cookie-based systems they make available to people online. There are many advantages of the header technique over the cookie-based technique; it is less complex and simple to locate and use, it is more persistent than cookie-based solutions, it addresses all forms of OBA-based tracking that may not all be cookie-based, and it does not rely on consumers finding, loading and managing lists of ad networks and advertisers to work.

However, it is important to point out that browser implementation of the DNT header does not represent a complete solution, as industry participation is required to create the technical mechanisms to respond to DNT browser requests broadcast by consumers via their browsers.

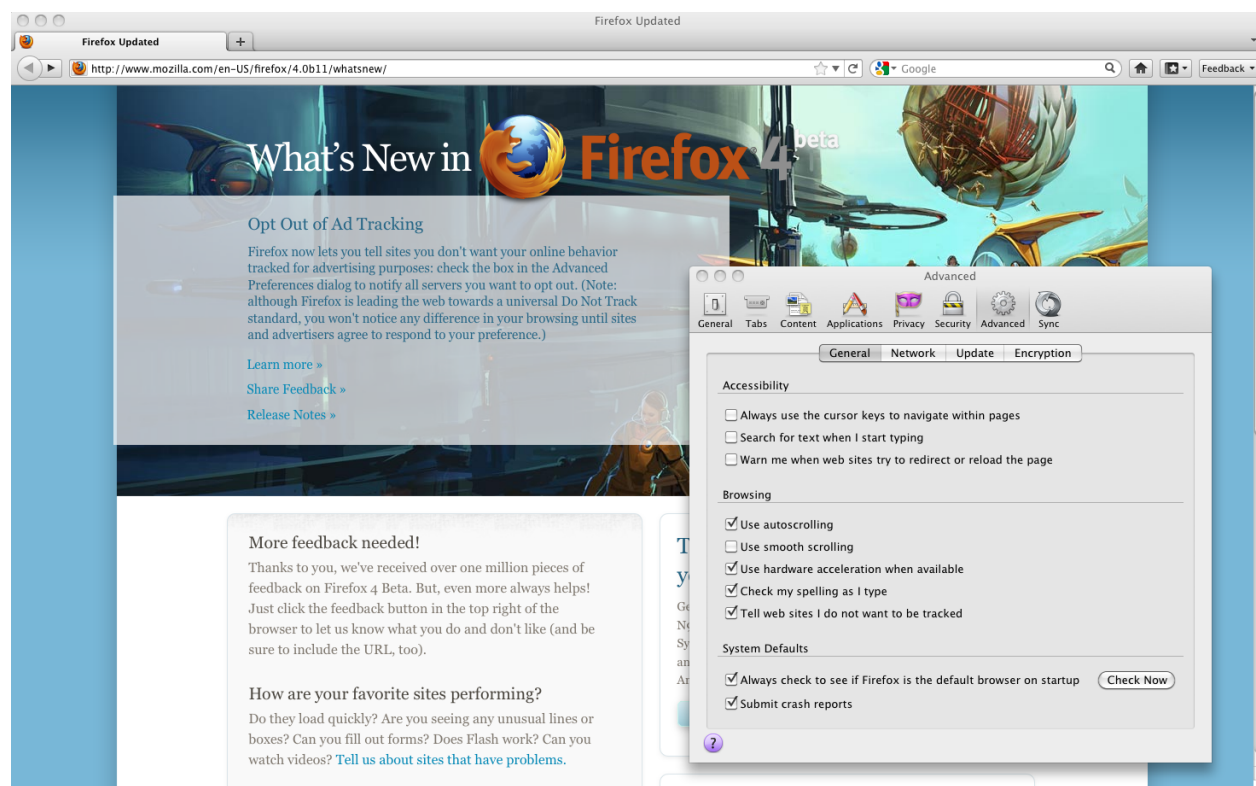
---

<sup>1</sup> "More Choice and Control Over Online Tracking," Alexander Fowler; <https://firstpersoncookie.wordpress.com/2011/01/23/more-choice-and-control-over-online-tracking/>

<sup>2</sup> "Opting-out of Behavioral Ads," Sid Stamm; <http://blog.sidstamm.com/2011/01/optiming-out-of-behavioral-ads.html>

<sup>3</sup> "Thoughts on Do-Not-Track," Michael Hanson; <http://www.open-mike.org/entry/thoughts-on-do-not-track>

## Screenshot: Firefox Welcome Page with Configuration Panel Open to Show DNT Header



Ad networks, advertisers and publishers are very supportive of the DNT header and see it as preferable to cookie-based or list blocking approaches. Consensus is emerging that a simple first step for responding to a consumer's intent could be: if the DNT header is present and the site or third-party advertiser has a tracking opt-out mechanism, then the mechanism should be activated. If the site or third-party advertiser does not have an explicit opt-out mechanism, the consumer should experience only content from a first-party relationship with the page being viewed. For behavioral advertising servers and data brokers, the intent of a DNT header is quite clear: it should be interpreted as though the consumer visited the opt-out registry and clicked the checkbox and that the consumer's activity or data is not collected or logged. We expect announcements to be forthcoming shortly on how first party and third party entities will be responding to the DNT header.

There are a number of steps ahead that will require continued leadership and support to see companies implement responses to consumers with the DNT header enabled, including:

- Fostering consensus on what the DNT header means to all stakeholders. We have proposed an initial definition focused on the display advertising market, and we seek a focused definition all stakeholders can agree upon.
- Helping to educate the public on DNT and what reasonable expectations of privacy people should have when using the DNT header or other mechanisms in a browser.
- Working with sites, advertisers and data brokers to establish best practices in implementing meaningful responses to a DNT header that are transparent to the public.
- Evaluating enforcement mechanisms to combat entities that systematically ignore the DNT header and jeopardize those efforts made by responsible companies.

## **Perspective on Working on DNT at the IETF and Tracking Protection Lists**

On March 7, 2011 we jointly submitted a draft proposal with Jonathan Mayer and Arvind Narayanan of Stanford's Center for Internet and Society to the IETF. The proposal, entitled "Do Not Track: A Universal Third-Party Web Tracking Opt Out," is a first attempt to define the syntax and semantics of a HTTP header-based mechanism for DNT and it also provides a recommendation for how web services should respond to such a mechanism.

At roughly the same time, Microsoft submitted a Tracking Protection proposal to the W3C containing three parts: Tracking Protection Lists (TPLs), the DNT header, and a doNotTrack DOM element. While TPLs provide a meaningful consumer protection for privacy, we do not think they necessarily fit well with the DNT header or DOM element; the goals and effects of the technologies seem to be quite different. For instance, TPLs affect how clients interpret and access content, while DNT header and the DOM element ultimately affect what servers do to preserve privacy. Additionally, there is no reason to limit deployment of the DNT header to web browsers; all HTTP-based communication could potentially benefit from this signal whether from a browser, application, or embedded device. We are in favor of moving the DNT header to a separate working group, preferably with the IETF, and then create a subcommittee of the W3C working group on TPLs to tackle standardization of the DNT DOM solution.

We recognize that the W3C has considerable experience working on privacy-related standards; however, HTTP is generally seen as the domain of the IETF. We also understand that the IETF may be a more open venue for stakeholders impacted by DNT headers who may not be members of the W3C, so that may be another factor to consider in selecting the appropriate venue.

### **About Mozilla and Privacy**

Mozilla is a global community of people working together since 1998 to build a better Internet. As a non-profit organization, we are dedicated to promoting openness, innovation, and opportunity online. Mozilla and its contributors make technologies for consumers and developers, including the Firefox web browser used by more than 400 million people worldwide. As a core principle, we believe that the Internet, as the most significant social and technological development of our time, is a precious public resource that must be improved and protected.

Privacy and security are important considerations for Mozilla. They are embraced in the products and services we create, and derive from a core belief that consumers should have the ability to maintain control over their entire web experience, including how their information is collected, used and shared with other parties. We strive to ensure privacy and security innovations support consumers in their everyday activities whether they are sharing information, conducting commercial transactions, engaging in social activities, or browsing the web.



SCHOOL OF INFORMATION  
102 SOUTH HALL # 4600  
BERKELEY, CALIFORNIA 94720-4600  
(510) 642-1464  
(510) 642-5814 Fax

March 25, 2011

## **Web Tracking and User Privacy position paper**

The W3C, the IETF, and other technical standard-setting bodies are poised to make a significant contribution to the development of scalable, technically-enabled approaches to privacy protection. Regulators, industry, advocates, and academics are looking to technical standards with renewed interest. The W3C should welcome the focus on the interplay between technical standards and social values and take this opportunity to fully enter the privacy conversation in a sustained and meaningful way.

### Collaborative and non-collaborative filtering

Effectively protecting user privacy in the face of ubiquitous and invisible tracking on the Web will likely require multiple policy and technical solutions. The experience dealing with unsolicited commercial email (spam) is instructive. Multiple technical and policy approaches were required to reduce the burden spam places on end users and networks. Spam filters allow for both black and white listing, while legislation and self-regulatory approaches that require labeling facilitate collaborative filtering.

The current proposals to address tracking for online behavioral advertising map these two approaches. Microsoft's member submission proposes a list-based blocking system, as well as a technical expression of a user preference (a Do Not Track header and property). The blocking operates much like black-lists in spam filtering, providing protection without the cooperation of other entities. This offers an important form of pre-emptive protection where the marketplace is comprised of entities with varying motivations to abide by users' wishes whether backed by law or not. The Do Not Track header/property, in contrast, requires that receiving entities abide by the expressed preference in order for privacy to be improved. Pursuing both options will allow Web browsers to work in both collaborative and non-collaborative settings, potentially improving privacy both in cases of good actors (who respect expressed user preferences) and bad actors (who might ignore or lie about their practices). Of course, as has been well-documented, the "arms race" of new tracking methods (HTTP cookies, Flash cookies, browser history sniffing, and on and on) suggests that tracking protection lists will not be the last necessary technical method for blocking tracking, nor need it be. In the same way, Do Not Track and other user privacy expressions may evolve beyond a single binary option.

### The need for a multi-stakeholder process

As illustrated by the technical proposals to address behavioral advertising, addressing privacy concerns requires coordination with non-technical parties and respect for the distinct spheres of expertise all participants bring to the discussion. The W3C's past experience with specifications within the technology and society domain suggest that a successful effort requires: 1) full participation of the

entities that must implement all aspects of the specification; 2) structures to maximize the ability of non-technical stakeholders with relevant privacy expertise to participate in appropriate elements of the specification; and, 3) participation that is geographically diverse to ensure technical interoperability despite competing policy approaches.

As with spam, the definition of the prescribed behavior — tracking — is not purely technical. Crafting the definition of tracking will require non-technical input. It may, as with the P3P vocabulary, argue for the creation of a separate expert group. Such an expert group should be broadly representative of the stakeholders and attentive to the need for responses that address varied global regulatory approaches. Technical approaches will be most useful if they support regional variations in privacy. A Do Not Track specification would be most useful if it interacts supportively with the ePrivacy Directive and opinions of the Article 29 Working Group as well as whatever regulatory and self-regulatory approaches emerge in the US and other countries. As in accessibility and P3P, precedent suggests that separating (but coordinating) technical and policy definitions can remove friction from the development process and leave flexibility where policy demands it.

### Technical standards and privacy by design

Focused work on the issue of behavioral advertising provides an opportunity to make an important contribution to a pressing public policy concern. However, privacy needs sustained attention. The current proposals to address behavioral advertising, like P3P before it, are episodic and largely reactive approaches to privacy.

The technical community has more to offer. Standard setting bodies have an important role to play in enabling privacy. Identifying approaches to the development of Web and Internet standards that provide sound building blocks for privacy protective designs, defaults, and policies requires a sustained and concerted effort. Equally importantly, the call for privacy considerations to inform design should not be exclusively led or dictated by lawyers or regulators. The effort must be a partnership. Identifying approaches to build privacy in will require active engagement between computer scientists and engineers, and privacy experts from other disciplines.

Privacy, like security, should yield a set of technical properties that can be defined and realized in various parts of the ecosystem. The properties that privacy may drive at the level of Internet or Web standards may be quite thin—in fact they may be properties that promote a broad set of policies. For example, properties of transparency, the ability to associate rules to data, and user control would provide hooks for privacy as well as other values (accessibility, choice and competition, for example).

Regardless of the ultimate outcome of the web tracking activities under consideration, the W3C should continue to expand its work on privacy. The W3C is uniquely positioned to sort out the appropriate role for Web standards in facilitating privacy solutions and has institutional experience building the bridges between disparate communities that is required to do this work.

Sincerely,

A handwritten signature in black ink, reading "Deirdre K. Mulligan". The signature is fluid and cursive, with the first name being the most prominent.

Deirdre K. Mulligan

## **Identifiers and Online Tracking**

Submission to W3C Workshop on Web Tracking and User Privacy

April 28-29 2011, Princeton, NJ, USA

Ashkan Soltani, Independent Researcher and Consultant

March 24, 2011

### **I. INTRODUCTION**

In the active discussions around ‘Do Not Track’, there seems to be some debate around what constitutes “tracking” and what consumer should expect when they signal that they do not want to be tracked online.

Proposed definitions for opting out of tracking range from companies agreeing to not collect or retain information resulting from online interaction to more obtuse definitions such as not serving personalized ads to these users, but still allowing for data collection. Some definitions provide stronger privacy protections to consumers (albeit with potential burdensome technical requirements for the ad networks) while others will still enable companies to collect (and perhaps even monetize) users’ data, even if they have indicated they don’t want to be tracked.

*In this short position paper, I propose a potential alternative approach to framing tracking that enables companies to engage in measurable online advertisement while providing the most important privacy protections articulated by advocates. This approach focuses primarily on the active removal of persistent identifiers that are used to correlate browsing activity over multiple sessions or multiple websites.*

### **2. CURRENT DEFINITIONS OF TRACKING**

There are various definitions what it means to opt-out of ‘tracking’ but I will summarize them into to primary camps:

#### **A. Do Not Track = Do Not Use For Behavioral Advertising**

Under the current system of ad network opt out cookies, consumers can opt out of the use of their data. That is, when a user ‘opts-out’, companies continue to track the consumer and even build a profile. However, they pledge to not to use this information for targeting although little is known about the secondary uses of this data, such as resale to other companies. This is the least privacy preserving option and arguably even a worse outcome for consumers since they have even less visibility to the data collection that is occurring yet do not receive the benefit of relevant ads. They still pay the privacy “cost”, but receive none of the benefits.

#### **B. Do Not Track = Do Not Collect or Retain**

Others are pushing for a Do Not Track system requiring companies theoretically delete all information received through third party transactions from consumers indicating that they do not wish to be tracked.

While this would certainly ensure that no private data would be stored by the third party, this implementation is has been criticized by website operators as being overly burdensome or difficult to implement. In order to comply with this definition in the **strictest** sense, they would be required to potentially configure all of the networking equipment and web servers they operate to not log data or delete it immediately. Load balancers, networking switches, routers and SSL accelerators would potentially all need to be modified to 'respect' the header and not log the browser request since most network infrastructure is built to log requests by default.

Furthermore, definitions in this category carve out multiple exceptions that allow collection and retention of data for specific uses, such as proving security, verification of ad impressions, or fraud detection. These exceptions will likely need to be crafted carefully and updated frequently in order to allow site operators to reliably serve content and innovate while still adhering to what most consumers expect when they request to 'not be tracked'.

### **3. DO NOT TRACK = DO NOT IDENTIFY?**

Much of the third party tracking that occurs online hinges on the presence of unique persistent identifiers which allow 'trackers' to identify individual users or devices across multiple visits to the same or different websites. These identifiers can be of the form of browser cookies although recent advances have given ways to other methods of identification, such as device fingerprinting and/or persistent storage outside of a browser's direct control.

Under this proposal, companies that agree to respect the Do Not Track signal could voluntarily make a best faith effort to strip **any** unique identifiers associated with the user/browser/client device as part of a web transaction after the transaction has occurred. The remaining data can be retained assuming that it doesn't later prove to be identifiable based on existing 'best practices' in identification.

This approach is good for business and consumers as it would allow businesses to collect and use data about how their websites are being used while preventing the creation of profiles. Fewer exemptions would need to be created since traffic management, fraud detection, and verification of impressions could all occur without relying on the uniquely identification of a individual device or browser persistently.

Much like a secret ballot: everyone gets the benefit of voting and the votes are tallied accurately, but no one can tell who voted for whom.

### **4. POTENTIAL IMPLEMENTATION**

Third party tracking consists of 3 key components, present in nearly every connection your

browser makes:

OBSERVER: the third party site that is tracking your activity

IDENTIFIER: unique descriptors that allow the 3rd party site to uniquely track you

ACTIVITY: i.e the URL of the 1st party site you're viewing (often the referrer url)

Consider the following snippet of data generated by viewing a page on the **WashingtonPost.com** about **insulin** which included a third party advertisement from **Mediaplex.com**:

**observer**      **identifier**      **activity**

```
http://img-cdn.mediaplex.com/0/16399/123458/Sms_30Umbrella_Definition_300x250.jpg
GET /0/16399/123458/Sms_30Umbrella_Definition_300x250.jpg HTTP/1.1
Host: img-cdn.mediaplex.com
User-Agent: Mozilla/5.0 (Macintosh; U; Intel Mac OS X 10.6; en-US; rv:1.9.2.16) Gecko/20110319 Firefox/3.6.16
Accept: image/png,image/*;q=0.8,*/*;q=0.5
Accept-Language: en-us,en;q=0.5
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 115
Connection: keep-alive
Referer: http://www.washingtonpost.com/national/insulin-is-key-in-both-forms-of-diabetes/2011/03/09/ABhN6D8\_story.html
Cookie: svid=192775639468; mojo3=16399:2151
```

In this request, the **observer** *img-cdn.mediaplex.com* is able to observe that a browser with cookie **identifier** *svid=192775639468* viewed the page *insulin and diabetes* page (**activity**) on the *washingtonpost.com* website.

Upon repeat activity, the **img-cdn.mediaplex.com** can correlate multiple visits of *washingtonpost.com* into a browsing profile keyed off of their cookie: **svid=192775639468**. if the user visits other websites which display third party advertising from **img-cdn.mediaplex.com**, then Mediaplex can correlate this activity across these sites as well, based on the same unique cookie id.

#### A. Do Not Use For Behavioral Advertising

Currently, some third party trackers allow the user to opt-out of tracking. However, this definition of 'opt-out' varies from third party to third party. While some websites allow users to opt-out of tracking by deleting or masking their cookies they still are able to identify users based on other factors such as IP address or Flash cookies.

In the above example, Mediaplex may allow a user to delete their **svid** cookie, but is still capable of profiling them based on other identifiers such as IP address or browser fingerprint.



While not typically apparent to the user, Mediaplex's systems would have an internal identifier they are utilizing.

#### B. Do Not Collect or Retain

Conversely, if this user wants to opt-out of this tracking completely, based on the 'Do Not Collect or Retain' definition, we could require that Mediaplex delete all of the log and profile data associated with the web request above. This is effective for consumers that don't wish to be tracked, but would likely make it difficult for Mediaplex to keep a record of this ad impression for accounting purposes.

As such, multiple exemptions may need to be created to allow third parties to retain browsing information in order to provide their basis accounting and security which ultimately goes against what consumers may expect when they believe they're not being 'tracked'.

#### C. Do Not Identify

Instead of asking Mediaplex to log no data at all, we could potentially request that third party websites strip *any persistent unique identifiers* from requests from consumers indicating that they do not wish to be tracked. In this case, that would mean stripping the unique cookie id although it could mean stripping other identifiers if they occurred in other portions of the request, such as the URL or Referrer header. If other identifiers, such as a browser fingerprint, are utilized on the back-end, the company would also be required to remove these as well. This 'stripping' can occur immediately or after a reasonable amount of time (i.e 24hrs) to facilitate processing of the transaction, though this is something that still needs to be worked through.

This approach would allow websites to collect information for the purpose of ad impressions, anti-fraud, security, and other purposes that are not user-specific. In fact, this is actually the current practice of many big advertisers who delete identifiers in log data as a result of that interaction, including but not limited to IP address, cookies, referrers, etc.

Participating websites could make a good faith effort to employ best practices in de-identification of their data based on evolving research in the field. Since these websites are the ones typically creating these unique persistent identifiers, they are in the best position to determine which information needs to be removed in order to make the data impervious to profiling.

#### D. Added Benefit for Monitoring Compliance

While all of these definitions require participation by website operators, the 'Do Not Identify' approach has the added benefit that allows web browsers or browser extensions to monitor web traffic and help identify any unique identifiers, such as cookies or URL parameters that are embedded in the content from sites that the user has signaled 'Do Not Track'. This could

indicate that this company may be engaging in unauthorized or accidental tracking. Browser fingerprinting or obfuscated identifiers are obviously still possible by rogue trackers, but this issue exists in the 'do not track = do not collect/retain' context too, i.e we're trusting the 3rd party websites to actually comply.

#### E. Mixed First/Third Party Interactions

Finally, this approach allows companies that operate simultaneously in first and third party context to comply with Do Not Track with no significant advantage to those that simply have third party presence, such as traditional ad networks. Companies that the user has a first-party relationship with, such as social networks or video sites, would still be able to serve personalized third-party content, such as social widget or 'over 18' video content, as per normal based on the identifier that was created during a first-party visit. However, the third party social network or video site could still be required to strip any unique identifiers in the subsequent tracking data recorded by these passive third party impressions.

Once the user takes direct action with a third party object, such as clicking a 'Share Widget' this could potentially convert the interaction into a first party user experience and fall outside of the scope of Do Not Track. However 'forced' first party interactions, such as auto-playing of an embedded third party video that the user must dismiss with should still be covered.

#### 5. CONCLUSION

While much discussion and clarification is needed to properly define what companies should do to comply with Do Not Track, focusing on identifiers *could be* a simple approach to reducing unwanted tracking/profiling while still enabling companies to engage in measurable online advertisement. Companies in the space can then innovate on ways to provide ads and services in a reliable way that does not infringe on a users desire to not be tracked/profiled.

Harlan Yu<sup>1</sup>  
Department of Computer Science  
Center for Information Technology Policy  
Princeton University

W3C Workshop on Web Privacy and User Privacy  
April 28-29, 2011

## **Accurately Communicating the Do Not Track User Preference**

One advantage of the *user preference* approach to Do Not Track is that users don't need to know in advance whether servers engage in tracking activities. The user simply needs to communicate her preference to the server, and the burden will then be on the server to refrain from any tracking. A simple and elegant way to communicate the user preference is using an HTTP header. Most users can choose a *blanket* tracking preference—to always send an “enabled” header to all sites or to never send the header to any site.

However, some users may decide to make more fine-grained tracking choices. For instance, a user could signal a preference to not be tracked only to third party domains and not to first party domains. Or, the user could consent to tracking by some third party domains and not others. Or further, the user could consent to tracking by some third party domains only when they're present on certain first party websites—while signaling to all other first and third parties a preference not to be tracked. The Abine Firefox extension has already made some of these finer-grained header preference options available to the user.<sup>2</sup>

In nearly all cases, the header should be sufficient to convey the user's tracking preferences. But, situations exist where the header may fail to accurately communicate the user preference, such as if a network intermediary unexpectedly strips the header out of the request. In other scenarios, the server may simply prefer to use an alternate technical mechanism to check the user preference. For example, a site using a complicated hosting infrastructure may find it easier to detect the user's preference using client-side code, rather than at the server that initially receives the HTTP request.<sup>3</sup>

It may be useful for browsers to include a client-side hook, accessible via Javascript, which conveys the same user preference as the header. The W3C submission from

---

<sup>1</sup> Email: harlanyu@cs.princeton.edu

<sup>2</sup> “To Track or Not to Track? Introducing DNT+.” Abine Privacy Blog, March 15, 2011. <http://abine.com/wordpress/http://abine.com/wordpress/2011/to-track-or-not-to-track-introducing-dnt/>

<sup>3</sup> Stamm, Sid. Comment on “DOM Flag” on the Do Not Track mailing list, March 14, 2011. <http://groups.google.com/group/do-not-track/msg/31df310ceb01c582>

Microsoft proposes the use of a DOM property for this purpose.<sup>4</sup> As currently proposed, the property is a global binary variable that is set uniformly for all domains. This is sufficient when the user has chosen a blanket tracking preference, but once a user decides to fine-tune her tracking preferences, the global DOM property will no longer accurately reflect the user's choice in every case.

One requirement of such a client-side mechanism should be that it *accurately mirrors* the user's original header preference. This means that the mechanism must support the same level of granularity as the header preference allows. An undesirable scenario is one where the HTTP header signifies an opt-out preference, but the DOM property misreports either opt-in or no stated preference. The server will have received a *conflicting* user preference, and the server may well proceed to track the user despite the header opt-out.

Some users will inevitably set more granular header choices. This is bound to happen, whether through functionality implemented directly in browsers or through extensions like Abine. It's not clear that DOM properties will be able to easily and accurately mirror the more fine-grained header choices.

To implement DOM access to user tracking preferences, a single DOM attribute such as `document.doNotTrack` will likely be insufficient. A better implementation would be an access method such as:

```
document.getTrackingPreference(in DOMString domain)
```

to look up the user's tracking preference for a domain from this document.

There is one significant implementation hurdle: *access control*. The problem is that when a first party site *includes* code from a third party, whether locally or remotely, the code will run on behalf of the first party, within the first party's protection domain. Thus, when client-side code calls the access method, the browser cannot tell which entity—the first party or a third party—is trying to access the information. This means, for example, that the New York Times (the first party site) could learn that the user consents to tracking by DoubleClick but not by Quantcast on its website. Moreover, even Doubleclick could potentially learn that the user does not consent to tracking by Quantcast from the New York Times site.

Resolving the access control issue seems quite challenging to overcome in today's browsers. Browsers don't currently *tag* the origin of client-side code. Even if it did, there would be no way for browsers to distinguish actual first party code from "third party code" that is added locally on the first party site. Poorly implemented access control could lead to a form of history-stealing attacks or make it even easier to fingerprint the browser. Additionally, if it is desirable that the user tracking

---

<sup>4</sup> "Web Tracking Protection." W3C Member Submission, February 24, 2011. <http://www.w3.org/Submission/2011/SUBM-web-tracking-protection-20110224/>

preference is *read-write* rather than read-only (as discussed below) these access control problems become even more pronounced.

### **Opting-back-in and Maintaining Tracking Transparency**

Another reason to potentially consider a client-side DNT mechanism is that servers may want to request that a user *opt-back-in* to tracking, inline in its Web application. A site may want to offer a special deal or premium services to an opted-out user, if the user is willing to opt-back-in to tracking. This could give sites a more flexible commercial framework to negotiate access to content or services, in exchange for tracking capabilities. Of course, any opt-back-in mechanism should carefully consider how to provide sufficient notice and obtain meaningful user consent.

But, regardless of whether a client-side mechanism is feasible, some sites may attempt to gain opt-back-in consent from users by storing the user preference server-side. Another undesirable scenario is where the user has selected the blanket preference to not be tracked, but certain entities continue to engage in tracking because the user—whether knowingly or not—has opted-back-in. Browsers would not be able to show in its interface which entities are still tracking the user.

As much as possible, tracking activities by servers should be transparent to the user. One potential remedy would be to implement a DNT *ack* header with the server's HTTP response. The ack would contain two parts. The first part just mirrors the DNT header from the HTTP request, so the user can verify that the preference was accurately received. The second part allows the server to report the user's tracking status.

For example, a DNT *ack* of "10" signifies two things. The "1" signifies that the server received a DNT:1 header in the user request. The "0" means that the server is still tracking the user, perhaps because the user has opted-back-in to tracking. Including an ack allows browsers to verify that DNT preferences are accurately received (and to notify the user when they are not) and to report in its interface how the user is being tracked, to the extent possible.

### **Separating the W3C submission on Web Tracking Protection**

On a separate note, the Microsoft W3C submission on Web Tracking Protection proposes two distinct technical concepts to deal with the same issue. The first approach uses filter lists to block certain unwanted user agent requests. The second approach describes a user preference for tracking to communicate user tracking preferences to Web servers.

While both approaches strive toward a similar goal, there's no reason why they need to be considered together from a technical perspective. It may be that users will find it most beneficial to adopt both technologies simultaneously, but it need

not be this way. Each approach has its distinct strengths and weaknesses, as well as separate technical and policy challenges. Indeed, browser vendors may decide to implement one approach but not the other.

I believe it would be beneficial to the general discussion around web tracking to separate these two approaches.

### **Acknowledgements**

Thanks to Ari Feldman at Princeton for helpful comments on this paper.