Do Not Beg:
Moving Beyond DNT through Privacy by Design

Mike Perry
The Tor Project, Inc
mikeperry@torproject.org

Abstract

The Do Not Track header (henceforth DNT:1) seeks to provide privacy protections against third party tracking through user request and regulation. It is our position that while DNT:1 is potentially useful as a purely informational tool for browser vendors and service providers, enforcement of the header suffers from a number of issues including covert circumvention, enforcement jurisdiction, manipulation, regulatory capture, and abuse. Moreover, every privacy property that DNT:1 aims to provide through regulatory enforcement can be better provided through technical changes to browser and network behavior during private browsing modes. We therefore suggest that the W3C standards body focus on standardizing these technical measures, rather than attempting to broker negotiations over regulatory policy and law.

1 Introduction

In this position paper, we describe the current and potential issues with DNT:1 and associated regulation, and also describe our prototype browser implementation[9] that aims to provide the same third party tracking resistance properties as DNT:1, but without relying on costly regulation and auditing.

We also believe that third party privacy can become a competitive feature for browser vendors, Internet service providers, and privacy preserving overlay networks.

2 Overview of DNT:1

The Do Not Track header seeks to provide users with a uniform mechanism to opt-out of third party tracking. Third party content elements are supposed to honor the header by declining to set cookies and record user activity on their servers. The draft standard[6] states that first party sites do not need to alter behavior with respect to the header. It also states a number of exceptions where third parties may still choose to retain and analyze data.

2.1 Benefits of DNT:1

The primary benefit of the Do Not Track header is that it provides a strong signal to browser vendors and websites with respect to their user's interest in privacy. Within a few months of the header's appearance, 7% of desktop and 18% of mobile Firefox users dug through the Firefox privacy settings to enable it.[4]

However, despite the value of sizing the market segment for frictionless privacy enhancing web technologies, it is very likely that DNT:1 will become a total disaster once the transition to regulatory enforcement draws near.

2.2 Shortcomings and Dangers of DNT:1

The primary shortcoming of DNT:1 is that it in no way alters the behavior of numerous browser technologies that enable and facilitate third party tracking, and instead relies entirely on ad-hoc auditing and potentially even regulatory enforcement.

Should strict auditing and direct regulatory requirements be enforced in some jurisdictions, it is very likely that at least some portion of the advertising industry would relocate to more favorable jurisdictions. In fact, they would be incentivized to do so, since it would allow them to offer advertising services at more favorable rates than their competitors who do not.

Similarly, it introduces serious risks of regulatory capture, especially in jurisdictions where advertising is able to wield considerable political influence over the selection of elected officials.

To answer these concerns, some DNT:1 advocates claim they favor "Carrot and Stick" incentive schemes that do not involve direct regulation, but instead will rely on web crawls to determine suspicious third party activity[7]. Their claim is that violators can be added to an always-on adblocker filter, and good actors could even be given immunity from data breach notification requirements and related privacy regulations.

However, without changes to the underlying browser technologies, there are simply too many ways for advertisers to covertly encode identifying information in third party elements. Even seemingly innocuous changes such as minor Javascript and CSS alterations across multiple elements can be used to encode covert identifiers that are stored in the browser cache for use as third party tracking cookies. This doesn't even begin to scratch the surface of covert third party identifier storage and supercookie vectors, let alone IP address and fingerprinting-based vectors, all of which we will discuss in more detail in later sections.

Further, behavioral targeting can be made very subtle, and difficult to distinguish from random chance. For example, Target has begun taking great pains to obscure behavioral targeting in its catalogs, to avoid alienating customers. Their targeted advertisements are still present, but they are merely blended with off-target messaging to provide a false sense of security and privacy[2]. It is extremely likely that such techniques will be employed by bad actors in the third party advertising world as well.

Further still, because of the various exemptions allowed in the DNT:1 standard, it is hard for users to know when the header is being honored, and if their activity is still being recorded, exchanged, and sold.

2.3 Hidden Costs of DNT:1

We believe that DNT:1 has seen such favorable adoption by browser vendors because of the ease of deployment for them. Adding a single HTTP header is substantially simpler than devoting research and development resources to addressing the network adversary in private browsing modes.

However, DNT:1 merely shifts the costs of privacy development and enforcement off of the browser vendors and onto every other party involved in the Internet economy, as well as onto new parties who were previously not involved (such as auditors, vigilantes, and governmental regulators).

Further, DNT:1 demands that standards organizations such as the W3C shift gears away from producing and reviewing technical standards to instead broker policy deals between regulators, legislators, and industry.

We believe that standards bodies and regulatory agencies shouldn't be wasting resources asking themselves how, when, and why advertisers don't obey DNT:1. Instead, they should be asking themselves why browser vendors whose revenue streams are often directly related to advertising markets continue to deploy technologies that facilitate and encourage covert third party tracking with no technical alternatives, even when their users enable their so-called "private browsing modes".

3 Do Not Track through Privacy By Design

Remarkably, the very same third party tracking resistance properties suggested by the DNT:1 draft standard are possible through a combination of browser and network behaviors.

All of these properties flow from a very simple core idea: two different first party domains should not be able to link or correlate activity by the same user, except with that user's explicit consent.

Initially, consent can be interpreted as link-click navigation. However, as federated login technologies such as web-send[1] and Persona[8] (formerly BrowserID) evolve, link-click based tracking vectors (such as the Referrer header) can be reduced or eliminated.

To understand the scope of the changes to the browser and network service providers to provide third party tracking resistance, we need to break down the problem into roughly four main areas of linkability and privacy: identifier sources, fingerprinting sources, disk activity, and IP address utilization.

3.1 Browser Behavior: Identifier Sources

Obviously, the primary vector through which third party tracking operates is the third party cookie.

Mozilla has a wonderful example of a first party isolation improvement written by Dan Witte and buried on their wiki[10]. It describes a new dual-keyed origin for cookies, so that cookies would only be transmitted if they matched both the top level origin and the third party origin involved in their creation. Thus, third party features could still function, if the user authenticated to that third party within the context of their first party url bar domain (perhaps using Mozilla's Persona, for example).

With respect to cache identifiers, the earliest relevant example of isolation work is SafeCache[5]. SafeCache eliminates the ability for 3rd party content elements to use the cache to store identifiers across first party domains. It does this by limiting the scope of the cache to the origin in the url bar origin. This has the effect that commonly sourced content elements are fetched and cached repeatedly, but this is the desired property. Each of these prevalent content elements can be crafted to include unique identifiers for each user, in order to track users who attempt to avoid tracking by clearing cookies.

Other identifier storage mechanisms that require such isolation include HTTP Auth, window.name, DOM Storage, IndexedDB, SPDY, HTTP-Keepalive, and cross-domain automated redirects. In Tor Browser[9], we either disable or isolate these technologies.

Properly isolating browser identifiers to the first party domain also has other advantages as well. With a clear distinction between 3rd party and first party cookies, the privacy settings window could have a user-intuitive way of representing the user's relationship with different origins, perhaps by using only the favicon of that top level origin to represent all of the browser state accumulated by that origin. The user could delete the entire set of browser state (cookies, cache, storage, cryptographic tokens, and even history) associated with a site simply by removing its favicon from their privacy info panel.

3.2 Browser Behavior: Fingerprinting Sources

After identifier isolation, the next source for covert tracking is through browser fingerprinting. Advertising networks can probe various browser properties known to differ widely in the userbase, thus constructing an identifier-free mechanism of tracking users.

Unfortunately, just about every browser property and functionality is a potential fingerprinting target. In order to properly address the network adversary on a technical level, we need a metric to measure linkability of the various browser properties that extend beyond any stored origin-related state.

The Panopticlick project by the EFF provides us with this metric[3]. The researchers conducted a survey of volunteers who were asked to visit an experiment page that harvested many of the above components. They then computed the Shannon Entropy of the resulting distribution of each of several key attributes to determine how many bits of identifying information each attribute provided.

While not perfect[1], this metric allows us to prioritize effort at components that have the most potential for linkability.

---

[1] In particular, we believe it is impossible to eliminate inter-browser fingerprinting vectors. Instead, fingerprinting metrics and defenses should focus on distinguishing features amongst a population with the same user agent. The Panopticlick test is not currently set up to do this.

This metric also indicates that it is beneficial for to standardize on implementations of fingerprinting resistance where possible. More implementations using the same defenses means more users with similar fingerprints, which means less entropy in the metric. It is for this reason (among others) that the Tor Project seeks to share its Firefox-based browser implementation[9] with any interested parties.

The fingerprinting defenses deployed by the Tor Browser include reporting the desktop resolution as the content window size, reporting a fixed set of of system colors, disabling plugins by default, limiting the number of fonts a document is allowed to load, and disallowing read access to the HTML5 canvas without permission.

The DNT:1 header itself is also fingerprinting vector for bad actors if we allow our users to set it, and the related scandal between Microsoft and Apache will likely cause us to entirely remove the DNT:1 option from Tor Browser's privacy preferences as a result.

3.3 Browser Behavior: Disk Activity

In addition to protecting against the network adversary, we believe that private browsing modes should not force the user to go without disk access. The two defenses are orthogonal, and private browsing mode users should still be allowed to store history, bookmarks, and even cookies and cache if they so choose.

Interestingly, a unified toplevel privacy UI could provide easy access to quickly clear all of these disk records on a per-site basis, using the same UI window for both tracking privacy and network storage.

3.4 Network Behavior: IP address utilization

Currently, there are many ways users can obtain a fresh IP address in an ad-hoc fashion. Users can use open wireless networks or tether to their phones. In fact, it is common practice for ISPs in many parts of the world to rotate user IP addresses daily, to discourage servers and to impede the spread of malware. This is especially true of cellular IP networks.

Obviously, only technically savvy users are likely to take full advantage of these properties correctly. However, there is no reason why an IP address allocation approach can't be generalized and standardized. One could imagine any privacy proxy (perhaps even one provided by your primary ISP) that intelligently isolates your first party page loads, along with all of their associated third party content, to a given IP address. By standardizing such a mechanism, privacy preserving networks can compete on network properties such as privacy or performance, rather than some combination of network and user agent.

The mechanism Tor has chosen to convey this information to the overlay network is the SOCKS username and password fields. Our plan is for the Tor Browser to inform the Tor client which network requests correspond to a given first party URL bar domain. The Tor client will then ensure that all first party loads use a different path through the Tor overlay network.

In fact, the Tor Project has concluded that it is in the best interests of the organization to share user agent development and standardization with other privacy preserving networks, both to reduce our development efforts, and to lead to a wider browser fingerprint population for our userbase. The German privacy company JonDos, GmbH has already joined this effort.

4 Conclusions

We discussed the Do Not Track header, the privacy properties it seeks to provide, and its shortcomings. We believe it is possible to provide these very same privacy properties through privacy by design.

While the DNT:1 header appears to be a simple change on the browser side, it has numerous hidden costs in terms of regulators, auditors, and server-side changes, in addition to serious regulatory challenges. We believe that it will actually be less costly in total to make the equivalent changes to the browser, and these changes will have the advantage of supporting markets for privacy proxies and related privacy enhancing technologies.

References

1. Tyler Close, Rajiv Makhijani, Mark Seaborn, Kenton Varda, Johan Apelqvist, Claes Nilsson, and Mike Hanson. Web Introducer. `http://web-send.org/introducer/`.
2. Charles Duhigg. How Companies Learn Your Secrets. `https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=9`.
3. Peter Eckersley. How unique is your web browser? In *Proceedings of the 10th international conference on Privacy enhancing technologies*, PETS'10, pages 1–18, Berlin, Heidelberg, 2010. Springer-Verlag.
4. Alex Fowler. Mozilla Led Effort for DNT Finds Broad Support. `https://blog.mozilla.org/privacy/2012/02/23/mozilla-led-effort-for-dnt-finds-broad-support/`.
5. Collin Jackson and Dan Boneh. Protecting browser state from web privacy attacks. In *In Proceedings of the International World Wide Web Conference*, pages 737–744, 2006.
6. J. Mayer, A. Narayanan, and S. Stamm. Do Not Track: A Universal Third-Party Web Tracking Opt Out. `https://tools.ietf.org/html/draft-mayer-do-not-track-00`.
7. Jonathan R. Mayer and John C. Mitchell. Third-Party Web Tracking: Policy and Technology. `https://www.stanford.edu/~jmayer/papers/trackingsurvey12.pdf`.
8. Mozilla Developer Network. Persona. `https://developer.mozilla.org/en-US/docs/persona`.
9. Mike Perry. The Design and Implementation of the Tor Browser. `https://www.torproject.org/projects/torbrowser/design/`.
10. Dan Witte. `https://wiki.mozilla.org/Thirdparty`.