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[1].

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).

---

[1] J. Mayer on twitter: "Another example of why domain names aren't the right privacy boundaries: metrics.apple.com = Adobe (formerly Omniture)"

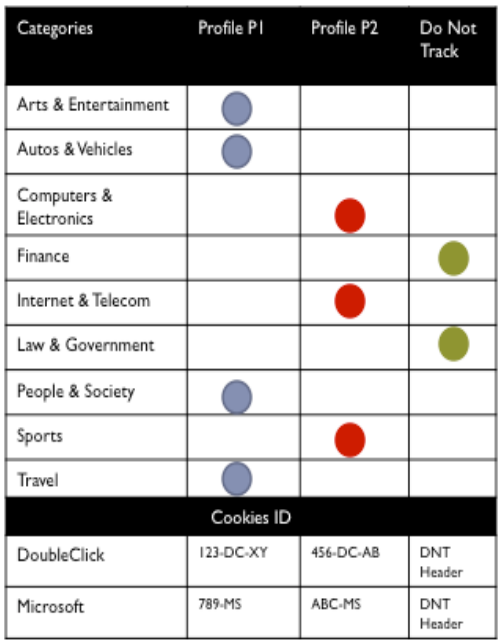

Figure 1: Profile configuration and mapping with cookies

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.

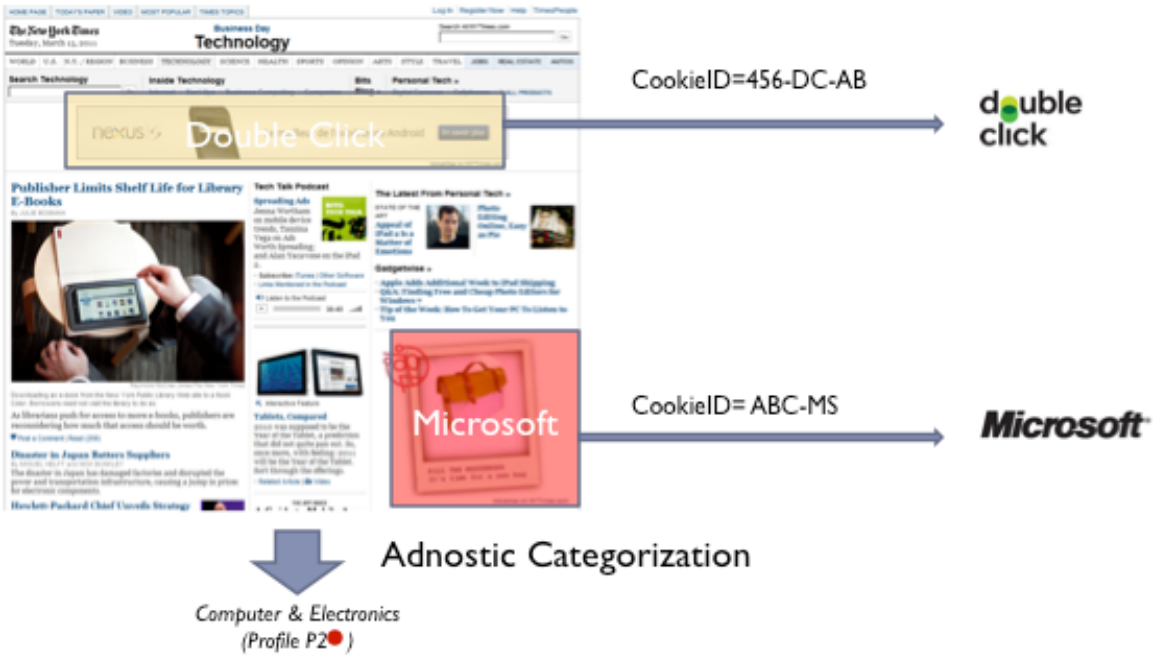

Figure 2 Pure browser solution description. The category is first determined (Computer and Electronics) and mapped to a profile. Then ads are downloaded with the right profile.

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/

[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