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@nsn.com. 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@blaeu.com.

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[1].

[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.

[1]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 Busineses 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 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.