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