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[1] [2], and Microsoft's Tracking Protection Lists (TPL)[3]. 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[4] on the history and lessons learned from the cookie specifications, including a substantial amount of history on

[1] http://dnt.mozilla.org/
[2] http://datatracker.ietf.org/doc/draft-mayer-do-not-track/
[3] http://ie.microsoft.com/testdrive/Browser/TrackingProtectionLists/
[4] 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[5], 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-

[5] 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?[6]

> *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"[7] online must automatically be considered a form of illegitimate tracking.

---

[6] 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.[8]

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".[9]

Unfortunately on the web, notions of ownership and contractual status are neither readily apparent nor manifest in the DNS. As such, it 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.

---

[7] http://en.wikipedia.org/wiki/Mashup_(web_application_hybrid)

[8] http://www.huntonfiles.com/files/webupload/CIPL_Use_and_Obligations_White_Paper.pdf

[9] http://en.wikipedia.org/wiki/Third_party

- www.paypal.com and www.paypalobjects.com are both operated by PayPal and both domains are integral to the operation of the 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[10]. Self-evident, if it was as easy as building in some technical controls, we wouldn't be having this discussion.

[10] http://ftc.gov/os/2010/12/101201privacyreport.pdf