APIs and consumers' privacy decision-making

– position paper –

Sören Preibusch ⟨Soeren.Preibusch@cl.cam.ac.uk⟩

University of Cambridge, Computer Laboratory

Cambridge CB3 0FD, UK

1 Privacy policies and how they do not guide privacy decisions

For more than a decade, service providers have posted privacy policies on their Web sites. These are intended to inform potential users of such sites about the service provider's practices in handling personal information. Users could thus read the privacy policy and make up their mind whether they are happy with the policy and willing to release their data or whether they feel uncomfortable with the practices described. Users may also find that data collection is excessive and the service provider is asking for far-reaching usage rights, but still, good service quality and low prices may compensate for that loss in privacy.

This architecture assumes rational consumers perform a utility-maximising calculus across all market alternatives by weighing the pros and cons of releasing personal information in the light of the benefits and risks it entails. Bounded rationality has made this approach fail in providing better privacy protection. The interaction or presentation layer is one of the pitfalls: on the one hand, textual privacy policies have been found to be vague, difficult and time-consuming to read and comprehend, technically inaccessible, lacking in fundamental details. On the other hand, the visual presentation of benefits in diligently filling out Web forms (such as qualifying for discounts or special announcements), is prominent. As a result, users find themselves giving up privacy unknowingly or even unwillingly, i.e. without informed consent.

P3P was designed as a technical approach to encode privacy policies so that decision support tools could parse a Web site's privacy policy and alert the human in front of the browser in case the Web site was exhibiting unwanted or potentially dangerous practices. The consumer would no longer have to read textual privacy notices. By lowering the cognitive burden in making privacy decisions, P3P would be a mechanism to cope with bounded rationality. Unfortunately, the adoption of P3P has remained low and published P3P policies are often malformed.

More recently, large platform providers, most prominently online social networks, have started to embed third-party supplied applications into their own Web sites. External content that comes with its own privacy policy is nothing new—P3P was designed to cope with the early forms, such as images and cookies delivered

by advertising networks, later Web analytics and social bookmarking. However, when applets are executed in the context of authenticated users, through APIs, these third-party Web applications gain read/write access to strongly identifiable personal information. Effectively, this corresponds to tiered privacy policies. Current practice is to make users consent upfront to grant the hosted applet full access; this carte blanche-access covers a subset of personal data which the original platform operator determines. Consent is typically agnostic of the applet's specific data requirements.

2 APIs for incremental informed consent and privacy-enhanced architectures

Application programming interfaces (APIs) to manipulate information in the underlying system now reverse the problem of privacy communication: they enable presentational diversity on top of the same underlying data protection semantics. A service provider that provides an API in addition to its own genuine Web form, enables third party software developers to build diverse presentational layers on top of its API. These third-party applications inherit the privacy policy of the API but may provide different clues to understand the amount and kind of personal information that is collected. Second, once an API is exposed, functionality can easily be wrapped inside a privacy-enhancing architecture. For instance, external applications guard endpoints of data flow and apply data encryption so that personal information may even be hidden from the original service provider. Third, APIs may potentially expose rich semantics, such as a WSDL document describing a Web service API. These are absent in the Web form the consumer normally faces. For instance, an API typically encodes which parameter values are required and which are optional whereas a Web form displayed in the browser relies on proprietary visual clues to mark mandatory input fields (e.g. through starring). Once the service specification exposed through the API also encompasses privacy parameters, these can be mined and alternative services may be compared and ranked by their privacy design.

Whilst APIs have acquired a reputation of leaking personal information out of Web companies' databases, they could be turned into a privacy enhancing technology or support the latter by sparkling competition on privacy: browsers, unifying add-ons and standalone applications as well as privacy-aware aggregators and search engines would interpret and emphasise differences in privacy designs and hence reward privacy-friendly service providers. We think that the 'required' attribute on input elements as well as their data-typing in HTML5 is a light-weight and yet promising step in this direction. Encoding alternative data inputs (e.g. home address vs. office address), potentially grouped by existing constructs such as fieldsets, would yield even more choice.

As soon as advanced APIs incorporate feedback mechanisms, bulk authorisation for data retrieval can be replaced with fine-grained access control. In the case of applets running inside a social network where

they interact with the network operator's API, an applet could signal which data items it actually requires. Eventually, the user would only need to give her consent to this limited subset of data items. Different versions of an applet could be tailored to have differing data requirements. An applet could dynamically extend its data coverage as the user requests higher service levels and gives consent in marginal increments. The concept of limited retrieval is already supported by some APIs, for instance through 'ResponseGroups' in interfaces Amazon provides. What is new, is that in a social networking usage scenario, the user's delegated powers to share data would need to be enforced by the operator's back-end processes. Advanced APIs therefore provide a deployment opportunity for privacy negotiations, by which users and service providers establish, maintain, and refine privacy policies as individualised agreements. The institutionalised framework of a single platform operator promises a short- to mid-term deployment perspective.

3 Advanced APIs and successful competition on privacy – workshop contribution

In our research, we study the economics of privacy. Our focus is commercial viability of privacy-enhanced service architectures. We investigate how consumers make privacy choices on competitive markets when data protection is not the only issue at stake. For instance, when consumers can choose between a privacy-friendly vendor and an alternative vendor that is marginally cheaper but requests additional data items, a plurality of consumers choose the latter. Other research has shown that phrasing and framing of data collection and salience of protective measures can reverse consumers' behaviour from reluctance to eagerness in disclosing items of personal data. This poses the question as to whether and how APIs will bring diversity into privacy configuration interfaces. Currently, manually restricting use of one's personal information often becomes a difficult or cumbersome task. Eventually, community applications built on top of highly popular APIs (notably in the areas of online shopping and online social networks) could bring a shift in data protection and allow users to exercise their rights more rigorously.

One lesson we have learnt in our laboratories is this: just because a service offers better privacy does not mean it will get used, let alone valued. This finding even holds if two services are fully identical but by their privacy policy. As long as we still have a limited understanding into how consumers make choices on the Web when it comes to privacy, we are unable to fully appreciate which privacy enablers need to be built into APIs. Consequently, we are still in an early learning phase how companies can monetarise good privacy practices.

In participating in the Workshop on Privacy for Advanced Web APIs, we want to take the opportunity to emphasise that introducing and liberally exposing APIs has an impact on behavioural and economic aspects of consumer privacy.

We want to share our experience in how researchers and practitioners can assess the business impact of privacy-friendly practices through carefully deployed user studies. As shown above, powerful APIs enable the deployment of privacy-enhanced services as well as the privacy-enhancement of existing services. To the extent to which users have a willingness to pay for such services, why should companies not internalise these revenues? We believe that service providers do not offer enhanced data protection as a measure of goodwill; on the contrary, data protection has the reputation of being a business impediment as it restricts the ability to extract the commercial value that resides in personal information. Yet, our empirical evidence suggests that being on the forefront of privacy protection can have a positive business impact.

We are looking forward to sharing insights as to how communication of privacy practices influences consumers' choice; how advanced APIs enable institutionally-sustained privacy negotiations and how they allow third-parties to monetarise data protection; and we are interested to learn more from the operators' perspective.