# A Case for Addressing Privacy Problems with Technical, not Legislative, Solutions

*Privacy Roundtables – Comment, Project No. P095416*

Randy Baden, Adam Bender, Bobby Bhattacharjee, Dave Levin, Neil Spring
*University of Maryland*

## 1 Introduction

Free online services have become an indelible part of today's Internet usage experience. The online experience of many users centers around web search (Google, Bing), web-based email (Gmail, Hotmail), online social networks (Facebook, MySpace), and video sharing (YouTube). Each of these services demand that users share their personal or usage information to some extent. Web-based email providers like Gmail, for example, require the ability to read users' emails so that they may target advertisements based on an email's contents. Ostensibly, other reasons for which services wish to have access to user data include simply making the service better, debugging and maintaining the service, or personalizing the user experience.

Unfortunately, users are often unaware that they are making their private information available to a company they have no compelling reason to trust. Further, many important questions remain unclear at best: Once a user uploads their data, who owns it? What does the service do with that data? What steps does the service take to ensure the confidentiality and availability of the data?

We believe that user privacy is paramount in today's Internet. Provable privacy is not only good for users of a service, it is good for the service itself. For instance, if a service never stores users' information unencrypted, then that service is indemnified from attacks on their own infrastructure. After all, even the most earnest security policies can be undermined by those who have physical access to the machines storing the data, such as disgruntled or careless employees. Maintaining user privacy also opens opportunities for more to use the service than could before. For example, many universities cannot move their email service to a web-based system due to privacy policy restrictions, even though doing so could lead to extensive savings. Were email providers to ensure the provable privacy of the colleges' data, this roadblock would be removed and web-mail providers might profit from ads targeted to the new customers [4].

We argue that problems of user privacy have the potential to be addressed with technical solutions over legislative ones. We have found that, surprisingly, *provable user privacy and the powerful services that drive today's Internet are not mutually exclusive*, even assuming *no trusted third parties.* That is to say, users do not need to trust services like Gmail or Facebook with their private data, because these services do not even need read-access to personal data to provide the same functionality they do today. In the remainder of this document, we support this argument with two case studies from our research in privacy and networking.

We favor technical solutions that do not require trust and yet yield *provable*, not merely *mandated*, privacy. Primarily, we believe that economic and legal incentives to protect user data pale in comparison to making it impossible for untrusted parties to glean the data in the first place. Without provable privacy, we anticipate a *privacy gap* where new companies will be unable to get users to provide personal information in order to use a new, competing service. Instead, as privacy concerns grow, users will trust only the sites with which they have a long-standing relationship. In today's online services, adoption and retention require users' trust [5]. Reducing the necessary trust that a user must place in a service provider may lead to broader adoption of new services, and ultimately a more competitive environment.

The authors of this document are graduate students and professors at the University of Maryland, College Park. We collaborate extensively with researchers at AT&T Labs–Research, Microsoft Research, Bell Labs, and the Max Planck Institute for Software Systems (MPI-SWS). We study attacks on user privacy and ways to uphold privacy in the face of these attacks. Our research aims to provide as much user privacy as possible while minimizing required user interaction. Put another way, our goal is to protect users' privacy while providing the same (or better) services that they enjoy today.

## 2 Case Study: Online Social Networks

Online social networks (OSNs) have become a de facto portal for Internet access for millions of users. These networks help users share information with their friends. Along the way, however, users entrust the social network provider with such personal information as sexual preferences, political and religious views, phone numbers, occupations, identities of friends, and photographs. Although sites offer privacy controls that let users restrict how their data is viewed by other users, sites provide insufficient controls to restrict data sharing with corporate affiliates or third-party application developers.

Not only are there few controls to limit information disclosure, acceptable use policies require both that users provide accurate information and that users grant the provider the right to sell that information to others. Facebook is a representative example of a social network provider.

The Facebook "Statement of Rights and Responsibilities" [7] *requires* that users "not provide any false personal information on Facebook" and "keep [their] contact information accurate and up to date." Further, it states that users "grant [Facebook] a non-exclusive, transferable, sub-licensable, royalty-free, worldwide license to use any IP [Intellectual Property] content that [they] post on or in connection with Facebook."

## 2.1 The danger of losing user privacy

Members of OSNs face several privacy risks, of which they are often unaware. When users are prompted by OSNs to enter their personal details and thoughts, they happily oblige, thinking only of the friends that will have access to the information. However, many other entities have access to the data as well. Most obviously, the OSN must be trusted to not abuse its access to the data. Users implicitely trust the OSN to not reveal or leak the information to anyone that the user has not permitted. However, this trust is violated on a regular basis. Krishnamurthy and Wills have shown that by including ads and applications in OSN pages, third-parties hosting this content have access to user's profile pages [16]. These third-parties can match the name, photograph, and other personal information found on the profile page to cookies on the user's machine that track the user's browsing habits. They have also found instances of OSNs that leak information, such as email addresses, that their privacy policies specifically state will not be shared.

OSN users also face risks from other users. In July 2009, the wife of the head of British intelligence agency MI6 "caused a major security breach" by posting family photographs and details on Facebook [11]. As a result, the chief's cover was compromised and he was left open to blackmail. Employers have recently started examining employees' and applicants' OSN profiles, which has led to several high-profile dismissals [20].

## 2.2 Persona: A Privacy-Preserving OSN

To meet the privacy needs of an OSN, we created Persona [2], an OSN that puts privacy policy decisions in the hands of the users. Persona uses decentralized, persistent storage so that user data remains available in the system and so that users may choose with whom they store their information.

Through encryption, Persona allows users to store private data persistently with intermediaries, but does not require that users trust those intermediaries to keep private data secret. Modern web browsers can support the cryptographic operations needed to automatically encrypt and decrypt private data in Persona with plugins that intercept web pages to replace encrypted contents.

Although private data in Persona is encrypted, we were able to demonstrate that many OSN applications – such as the Facebook Wall, news feed, and profile – do not require access to data

contents, instead only requiring knowledge of the structural relationship between data items. For those applications that do require data contents (because local computation is either impractical or impossible), Persona users may optionally choose to selectively reveal information to applications exactly as they would choose to reveal it to friends.

Our initial evaluation of a Persona prototype indicated that Persona introduced overhead in exchange for privacy but that this overhead is small enough to realize a full deployment given current technology. Our microbenchmark on the iPhone indicated that Persona is even feasible on mobile devices.

## 2.3 OSN Impersonation and Privacy

OSNs have persuaded millions of users [8] to give their offline identities an online presence. While these OSN identities are convenient for online communication, they risk impersonation [10] and may provide personal information that threatens the security of other systems [17, 18]. Users, aware that their personal information is valuable, may choose only to allow their friends to see their information. However, even correct privacy settings can be foiled if someone has infiltrated their circle of friends. Users cannot trust that the person behind an online account is actually their offline friend, even if that account has the correct picture and profile information [3]. We therefore proposed a solution to the problem of OSN impersonation.

Offline we have ways of identifying a friend—such as recognizing her appearance—that are either difficult or impossible in online communication. When our sight is not sufficient to identify a friend, we might fall back on recognizing her voice, and if that fails, we fall back on other information that we know from past interaction with that friend. We proposed the use of *exclusive shared knowledge* for identification: we can identify a friend (both online and offline) by asking her questions that only she can answer.

Once we identify our friend, we can ask her to provide or verify a public encryption key associated with her identity. By repeating this process with all of our friends, we can bootstrap a public key infrastructure (PKI) that we can use on the OSN. This PKI is sufficient for detecting OSN impersonation and incorporating OSN identities into the wider realm of distributed systems research.

We performed a user study, Bond Breaker [1], that showed that when users employ exclusive shared knowledge strangers have less than a 2% chance of guessing the answers to shared knowledge questions; this compared favorably to web-based security questions—another identification scheme based on personal information—which can be guessed 17% of the time by strangers [19]. We show that even when users only exchange keys with a few friends, we can discover the keys

of many friends and friends-of-friends with a web of trust. Finally, we show that the same web of trust detects 80% of all successful impersonation attacks.

# 3 Privacy-Preserving Targeted Advertisements

Targeted advertisements fuel today's Internet. Virtually any website can obtain revenue from displaying advertisements, for example via Google AdSense [12]; Google reported earning over $21B in ad revenue in 2008 alone [13]. The efficacy of targeted advertisements is evident, and given the sheer volume of participating advertisers, it is safe to assume that this style of advertising is here to stay. Without a way to privately target ads, user privacy cannot be a reality on the Internet.

Our concern lies not with the existence of targeted advertisements, but with the privacy that users give up in order to be targeted. Not only is ad targeting a huge source of revenue for those who host ads, but we believe that it can improve a user's experience by presenting the information and products they wish to purchase. In the remainder of this section, we briefly describe what information ad-targeting sites obtain from users, as well as the properties of an architecture that we believe to be a viable replacement for today's scheme.

## 3.1 Information sharing in today's targeted ads

Because advertisers typically pay each time a user clicks on one of their ads, it is in the interest to ad-targeting sites like doubleclick to present ads that users are most likely to click on. To arrive at the right ad at the right time, ad-targeting sites analyze both a user's long-term, *behavioral* data (e.g., their search history over a long period of time) and short-term, *contextual* data (e.g., the content of the email they just opened). Put simply, the more an ad-targeting site knows about the user, the more precisely they can target ads, and hence the more revenue that site can generate.

## 3.2 Private yet accountable targeted ads

Over the past decade, there have been several proposed schemes to target ads while preserving user privacy [9, 14, 15]. The fundamental problem in providing privacy-preserving targeted advertising is in protecting against *click-fraud*. Broadly defined, a user commits click-fraud by clicking on an advertisement when they have no interest in purchasing that product. Advertisers have incentive to fraudulently click on their competitors' ads to reduce their ad budget, while website publishers have incentive to click on ads on their own page to increase their ad revenue. To be viable, a privacy-preserving ad-targeting system must therefore keep users' data and clicking behavior private while making their clicks *accountable* to advertisers.

Some of the authors of this document, along with colleagues from Microsoft Research—John Douceur, Jacob Lorch, James Mickens, and Thomas Moscibroda—have developed an architecture that targets ads while preserving user privacy.[1] The main properties of this architecture are as follows:

- *It makes ad targeting more private than ever, and yet able to access more information than ever.* Our design maintains user privacy with a similar approach to prior systems: by targeting the advertisements on the user's machine. Users' private data therefore does not have to be shared unencrypted with anyone else. An interesting side-effect of this security approach is that ads can be targeted based on even more information than possible today. In particular, there is no single online service that can track as much behavioral information as a user herself can.

- *It protects against click-fraud.* Even if users access the Internet behind an anonymizing proxy such as Tor [6], the architecture is able to detect when a user has clicked on a number of ads that would indicate click-fraud. This is a precise statement: an ad service provider can detect that *a* user has performed click-fraud, but cannot detect *which* user in particular. This upholds the necessary requirement of click-fraud detection without introducing potential security leaks if, for example, a user accidentally clicks repeatedly on a given ad.

- *It maintains today's economic model.* We believe that ad service providers would be reluctant to adopt a new scheme if it adversely affected their revenue stream. Our design maintains the flow of money—advertisers pay the ad service provider, and the ad service provider pays (to a lesser extent) the publishers. Crucially, our design does so *without introducing any additional parties, trusted or otherwise.*

To summarize, along with our collaborators at Microsoft Research, we have demonstrated an alternative to legislative approaches for protecting user privacy while targeting ads. We believe that approaches like this can simultaneously meet the economic needs of those *running* online services, while providing the privacy needs of those *using* such services.

# 4 Where do we go from here?

In this document and in our research, we have shown that *powerful online services and user privacy are not mutually exclusive.* That is, there are technical solutions to protecting user privacy

[1] This work is currently under submission to the USENIX Symposium on Networked Systems Design and Implementation (NSDI) 2010.

while maintaining a powerful service like OSNs or targeted advertisements. The next step is to begin adopting such technologies in today's services. We are not proposing the adoption of any particular form of technology. Rather, we are pointing out an alternative to today's approach of forcing users to share all of their information.

The privacy-preserving architectures we have discussed in this document have the property that *no one* can learn user information without the user having explicit knowledge and giving explicit permission. As we discussed at the beginning of this document, services benefit from personal information, for instance to understand how to improve their service. But this is not a new problem; the typical answer is simply to sample from users who opt in to releasing their private information. Nielsen ratings use statistical sampling of a small, voluntarily participating portion of the population to infer characteristics of television viewers as a whole. Medical patients must give full consent for doctors to publish their cases. These are but two examples that demonstrate that there are many technological solutions that obviate users sharing personal identifying information, particularly without full, explicit knowledge of what is being done to their data.

As a next step, we hope that:

- Policy makers consult researchers and practitioners to better understand what information does and does not *have* to be shared to provide valuable services in the Internet.
- Service providers provide full, explicit information on how users' data is being used, and with whom it is sharing that information.
- Users have a choice between privacy-preserving and open-book services. Perhaps, at the moment, users have some reason to trust the Amazons and the Googles who have been around a long time, but the Internet as a whole might benefit if we could (partially) trust a certification process or trusted hardware to adhere to privacy policies set by consumers, not by companies.

# References


[1] R. Baden, N. Spring, and B. Bhattacharjee. Identifying close friends on the internet. In *HotNets '09*, 2009.

[2] R. Baden, A. Bender, N. Spring, B. Bhattacharjee, and D. Starin. Persona: an online social network with user-defined privacy. *SIGCOMM Comput. Commun. Rev.*, 39(4):135–146, 2009.

[3] L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All your contacts are belong to us: automated identity theft attacks on social networks. In *WWW '09*. ACM, 2009.

[4] J. Caplan. Google and Microsoft: The Battle Over College E-Mail. TIME article, available at http://www.time.com/time/business/article/0,8599,1915112,00.html.

[5] K. Dearne. Google tries to allay privacy fears. Australian IT article, available at http://www.theaustralian.com.au/australian-it/google-tries-to-allay-privacy-fears/story-e6frgakx-1225794910369.

[6] R. Dingledine, N. Mathewson, and P. Syverson. Tor: The second-generation onion router. In *Proc. USENIX Security*, 2004.

[7] Facebook statement of rights and responsibilities. http://www.facebook.com/terms.php.

[8] Facebook Statistics. http://www.facebook.com/press/info.php?statistics.

[9] J. Freudiger, N. Vratonjic, and J.-P. Hubaux. Towards privacy-friendly online advertising. In *In Proc. of Web 2.0 Security & Privacy (W2SP)*, 2009.

[10] E. Friedman. Is Rick Astley Dead? Internet Hoaxes Have Fans Wondering. ABC News article, available at http://www.abcnews.go.com/Technology/Story?id=7960020.

[11] N. Gilani. Wife blows MI6 chiefs cover on Facebook. Times Online article, available at http://www.timesonline.co.uk/tol/news/uk/article6639521.ece.

[12] Google AdSense – Publisher Solutions. https://www.google.com/adsense/static/Publishertools.html.

[13] Google Investor Relations – Financial Tables. http://investor.google.com/fin_data.html.

[14] H. Haddadi, S. Guha, and P. Francis. Not all adware is badware: Towards privacy-aware advertising. In *IFIP Conference on e-Business, e-Services, and e-Society (I3E)*, 2009.

[15] A. Juels. Targeted advertising ... And privacy too. In *Proc. Conference on Topics in Cryptology: The Cryptographer's Track at RSA*, 2001.

[16] B. Krishnamurthy and C. Wills. On the leakage of personally identifiable information via online social networks. In *Workshop on Online Social Networks*, 2009.

[17] PC World. http://www.pcworld.com/article/168462/.

[18] A. Rabkin. Personal knowledge questions for fallback authentication: security questions in the era of facebook. In *SOUPS '08*. ACM, 2008.

[19] S. Schechter, A. J. B. Brush, and S. Egelman. It's no secret: Measuring the security and reliability of authentication via 'secret' questions. In *IEEE Symposium on Security and Privacy*, 2009.

[20] R. Stross. How to lose your job on your own time. New York Times article, available at http://www.nytimes.com/2007/12/30/business/30digi.html.