Access Control is an Inadequate Framework for Privacy Protection

Lalana Kagal and Hal Abelson
MIT Computer Science and Artificial Intelligence Lab
32 Vassar Street Cambridge MA 02139
{lkagal, hal}@csail.mit.edu

Introduction

As Web architects, we might all agree on the need to protect privacy. *But what is it that we want to protect?* In Brandeis and Warren's classic legal study [19], privacy is defined as the "right to be let alone". In Alan Westin's seminal work [20], privacy is the ability for people to determine for themselves "when, how, and to what extent, information about them is communicated to others". The UN Declaration of Human Rights [16] stipulates that "No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honour and reputation".

There are important differences among these definitions. One central definition is that Westin focuses on **information access**: how information comes to be known. In contrast, the UN Declaration, and Brandeis and Warren, focus on what happens to people as a result of **how information is used**.

Present discussions of Internet privacy, both policy and technology, tend to assume Westin's perspective. The focus on access control sometimes regards privacy as a kind of "currency of the digital age"[1] that people need to exchange in return for better search results, more personalization, customized services, more targeted advertising, and better communication with friends, family, and colleagues. "Protecting privacy" is often equated with letting users make these tradeoffs by defining detailed rules to govern access to their personal information. This year's technology press is filled with announcements by social networking sites about their new privacy controls, i.e. new ways for users to define access rules [18, 23]; followed by embarrassment when the choices prove to be inadequate or too complex for people to deal with [17, 1, 10, 22, 15, 3, 13].

Even when access control systems are successful in blocking out unwanted viewers, they are ineffective as privacy protection for a large, decentralized system like the World Wide Web, where it is easy to *copy* or *aggregate* information. These days, it is possible to infer sensitive information from publicly available information. For example, social security numbers (SSN) have always been closely guarded because they are used to identify individuals by most government and financial institutions. Every use

[1] As described by Yamini Kagal, June 2010

of SSN numbers is usually viewed as risking identity theft. However, CMU researchers recently showed how SSN numbers can be reconstructed with a high degree of accuracy from public sources including commercial databases and government sources [12]. *If SSN numbers can be inferred easily how does restricting access to them help our privacy ?* Social networking data can also be used to infer sensitive and possibly secret information about you. You may increase your privacy settings on a social networking site to prevent access to most of your data, however your friends or the groups you belong to may inadvertently say more about you than you'd like. Experiments have shown that it is possible to predict sexual orientation [7] and political affiliations [9] of users by simply looking at their social networks.

Another problem with using access control systems is that it is the responsibility of users to clearly define their privacy policies. Unfortunately users are often unable to comprehend the inherent consequences of their privacy policy. *Should you allow the friends of your friends to view your pictures ? What if your friend adds your boss as his friend, will your boss be able to see your party pictures ?* Studies show that most security failures can be contributed to user errors [21] caused by clumsy user interfaces for security. The same will be true about privacy failures if it is solely the user's responsibility to define privacy. Even if the privacy controls are simplified [23], it is infeasible to expect users to correctly define their policy or even understand what their policy should be. Users would also need to set these preferences for each of their sites/services, which would most probably use different terminology and controls making it even more complicated. Forcing users to make decisions about their privacy and expecting them to correctly specify and manage these policies across several systems is impractical.

The central claim of this paper is that access control in itself is inherently inadequate as a framework for addressing privacy on the Internet. We need to give more emphasis to Brandeis and Warren, and UN Declaration frameworks, and focus on how data is used, even more so than how data is accessed. There, in fact, may be little choice here. In a pure access restriction system, those who obtain access to the data, legitimately or not, can use the data without restriction.

We propose alternate strategies to access control, ones that are similar to how legal and social rules work in our societies. Vast majority of these rules are not enforced perfectly or automatically, yet most of us follow the majority of the rules because social systems built up over thousands of years encourage us to do so and often make compliance easier than violation. For example, when trying to find parking on the street even if there is no traffic police around to give me a fine or a ticket, I will still obey the parking signs because of the way social and legal norms have been built into our societies. They've made it easier to meet the norms than violate them, risk getting caught and getting punished. The set of legal or social norms govern what we can or cannot do, and they have been ingrained into our way of thinking such that most people go through life without any problems. When problems occur, there are mechanisms that ensure that violators are reprimanded. Instead of enforcing privacy policies through restricted access, we suggest focusing on helping users conform to policies by making them aware of the usage restrictions associated with the data and helping them understand the implications of their actions and of violating the policy, and encouraging transparency and accountability in how user data is collected and used.

In order to support the responsible use of private data, we believe information systems should be engineered to have the characteristics described below

- Give users due notice: Both in the case of collecting their data and using it, information systems must give users due notice. This will give users the opportunity to respond appropriately - to either take action to protect their privacy or voluntarily give it up in exchange for better service. Google AdSense provides a form of due notice by putting a hyperlink "Ads by Google" on all its advertisement. When clicked, the user gets general information about why these ads were displayed and is able to slightly modify how further targeting is performed. However, a better approach would be one that informs the user about exactly which information collected from her was used to choose the particular ad [6]. The Google dashboard [5] is another way that Google gives notice. The dashboard lets users know what data is being collected about them. Though it is a step in the right direction it does not inform users what the data is used for or what the consequences of misuse are.

- Don't rely just on upfront authorization, also support accountability: A-priori authorization provides a way to restrict access to sensitive data but has several problems as outlined in the section above. A supporting approach is to provide post-facto accountability, which requires the system to have mechanisms in place to identify misuse of data.

  - Provenance: It must be possible to track the data as it follows through the system and to maintain detailed provenance information. This information should be in a machine-understandable format so software tools can automatically reason over them to identify violations.
  - Machine-understandable policies: When systems give users due notice they provide policies outlining appropriate use of users' data. These policies can either be negotiated between the system and the user or be an interpretation of some laws that describe appropriate use of data such as HIPAA or the Privacy Act. Providing policies in a machine-understandable manner serves two purposes. They can be used by tools at the users' end to figure out if they meet the users' personal policies for data usage. They can also be used by mechanisms trying to identify inappropriate use in provenance trails kept by the system.
  - Policy tools: Systems should include policy tools that not only reason over policies and provenance to identify violations but also support day to day operations by answering questions about use of data such as *Can this data be used for a given purpose ? What are the consequences of misusing this data ?*

- Privacy-enabling interface design: As we want to encourage appropriate use not only by information systems but also by users, system interfaces should provide hints or signs describing optimal behavior for users with respect to data.

- Policy-awareness: Policy-awareness is a property of information systems that provides all participants with accessible and understandable views of the policies associated with information resources including the consequences of misuse. Creative Commons is a good example of providing useful policy-awareness [2]. It gives content creators the ability to specify different kinds of licenses and has created easy to recognize icons that encourage the proper re-use of content. Following the Creative Commons model, RespectMyPrivacy is a Facebook application that allows users to select their usage restrictions and displays it using icons on their Facebook profile [8]. Another MIT project extracts Creative Commons licenses associated with media and automatically provides appropriate attribution when content is re-used [14].
- Privacy implications: Users also need help to understand the consequences of their actions on the Web, some of which will have privacy implications. Some examples of these mechanisms include 'privacy nudges' and Google's Mail Goggles [4]. CMU researchers propose the notion of privacy nudges [11] where software agents monitor the users' keystrokes and inform them if they enter any personally identifying information on a Web form. Google's Mail Goggles attempts to prevent users from sending email they might later regret by asking them to solve a simple math problem before they can send an email. Other possible mechanisms include showing thumbnails of everyone your email is being sent to including members of a mailing list[2] and displaying a social network with images of everyone who would see the message or photo you're about to post. These mechanisms, though simplistic, help users understand where their information is going. By analogy with provenance, which deals with tracking where data comes from or the source of the data, this ability to understand where the data is going can be thought of as "abvenance"[3].

Discussion

Many of us take the right to privacy for granted, but it is not currently a part of the US Constitution. In some developed countries there are in fact federal laws protecting users against inappropriate data collection and use. However, in the U.S. we have a set of laws covering different aspects of information privacy such as the "do not call list" for telemarketing and HIPAA for medical record privacy. Unfortunately as technology constantly changes, these laws do not adequately protect users.

However, privacy cannot be solved by regulation alone. Education and technological solutions will also play an important role. Users should be educated on good privacy practices, be given mechanisms to understand the privacy implications of their actions and be encouraged to meet others privacy/usage restrictions. Technology will help privacy but will not solve the problem entirely. Using technology to decide fair

[2] As suggested by Tim Berners-Lee

[3] abvenance - Latin "ab" prefix: "off", "away from"

use has shown similar limitations leading us to believe that trying to develop a purely technical solution to privacy is also prone to failure.

As most online businesses today leverage user data to provide more customization, privacy can be seen as a tradeoff between value-added services and loss of personal data. As preventing access to personal data is no longer an option and as sensitive data can be easily inferred from public sources, we believe the future of privacy protection lies in ensuring the responsible use of data.

References

[1] 3 News. Australian PM caught following porn star. http://www.3news.co.nz/Australian-PM-caught-following-porn-star/tabid/412/articleID/135848/Default.aspx, 2010.

[2] C. Commons. Creative Commons licenses. http://creativecommons.org/.

[3] GigaOm. Is facebook beacon a privacy nightmare? http://gigaom.com/2007/11/06/facebook-beacon-privacy-issues/, 2007.

[4] Google Blog. New in labs: Stop sending mail you later regret. http://gmailblog.blogspot.com/2008/10/new-in-labs-stop-sending-mail-you-later.html, 2008.

[5] Google Blog. Transparency, choice and control now complete with a dashboard! http://googleblog.blogspot.com/2009/11/transparency-choice-and-control-now.html, 2009.

[6] S. Hansell. An Icon That Says They're Watching You. http://bits.blogs.nytimes.com/2009/03/19/an-icon-that-says-theyre-watching-you/, 2009.

[7] C. Jernigan and B. Mistree. Gaydar: Facebook friendships reveal sexual orientation. http://firstmonday.org/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/2611/2302, 2009.

[8] T. Kang and L. Kagal. Enabling privacy-awareness in social networks. In *Intelligent Information Privacy Management Symposium at the AAAI Spring Symposium 2010*, March 2010.

[9] J. Lindamood, R. Heatherly, M. Kantarcioglu, and B. Thuraisingham. Inferring private information using social network data. In *World Wide Web Conference poster paper*, 2009.

[10] MSNBC. Twitter gets you fired in 140 characters or less. http://www.msnbc.msn.com/id/29796962/, 2009.

[11] New York Times. Redrawing the route to online privacy. http://www.nytimes.com/2010/02/28/technology/internet/28unbox.html, 2010.

[12] PC World. Researchers expose security flaw in social security numbers. http://www.pcworld.com/article/167975/researchers_expose_security_flaw_in_social_security_numbers.html, 2009.

[13] PC World. Google buzz criticized for disclosing gmail contacts. http://www.pcworld.com/businesscenter/article/189081/google_buzz_criticized_for_disclosing_gmail_contacts.html?tk=rel_news, 2010.

[14] O. Seneviratne, L. Kagal, and T. Berners-Lee. Policy aware content reuse on the web. In *ISWC2009 - International Semantic Web Conference*, October 2009.

[15] The Local. Headmaster fired after Facebook pic scandal. http://www.thelocal.se/20148/20090618/, 2009.

[16] United Nations General Assembly. Universal Declaration of Human Rights (UDHR). http://www.un.org/en/documents/udhr/index.shtml, December 1948.

[17] UPI. Waitress fired for Facebook comment. http://www.upi.com/Odd_News/2010/05/17/Waitress-fired-for-Facebook-comment/UPI-39861274136251/, 2010.

[18] Wall Street Journal. Facebook grapples with privacy issues. http://online.wsj.com/article/SB10001424052748704912004575252723109845974.html?mod=WSJ_Tech_LEFTTopNews, 2010.

[19] S. D. Warren and L. D. Brandeis. The Right To Privacy. 4 Harvard Law Review 193, 1890.

[20] A. Westin. Privacy and freedom (Fifth ed.). New York, U.S.A.: Atheneum, 1968.

[21] A. Whitten and J. Tygar. Why johnny can't encrypt: a usability evaluation of pgp 5.0. In *SSYM'99: Proceedings of the 8th conference on USENIX Security Symposium*, Berkeley, CA, USA, 1999. USENIX Association.

[22] Wikipedia. Star wars kid. http://en.wikipedia.org/wiki/Star_Wars_Kid, 2002.

[23] Wired. Facebook debuts simplified privacy settings. http://www.wired.com/epicenter/2010/05/facebook-debuts-simplified-privacy-settings/, 2010.