

Prevention and Reaction: Defending Privacy in the Web 2.0

Michael Hart Rob Johnson
{mhart,rob}@cs.stonybrook.edu
Stony Brook University

Abstract

User data can originate from three sources – the user, the user’s friends and acquaintances, and services with which the user interacts – and users may be harmed by the disclosure of data from any of these sources, some of which are not directly under her control.

We argue that, for data disclosed by the user herself, tag-based privacy policy languages are flexible and easy for users to understand, enable users to express their preferences succinctly, offer good portability across service providers, and provide a strong foundation for creating sticky policies. We review our previous work developing and evaluating a tag-based privacy scheme. Our past results demonstrate that tag-based policies are easy to use, support short policies, and enable Web 2.0 services to offer automatic privacy policy inference, further reducing the policy maintenance burden on the user.

We then consider several possible mechanisms for managing the impact of data published about one user by another. Even if it were possible to automatically determine the target user described by the data, social conventions, such as freedom of expression, preclude giving the target user absolute control over the information. Furthermore, privacy and data usage controls must advance the interests of the parties that implement those controls, and this significantly constrains the design space for effective solutions. We therefore propose reactive mechanisms that users and search engines can use to steer other users away from libelous content.

1 Introduction

The reach of Web 2.0 applications has gone beyond most everyone’s expectations. Facebook started as a social network confined to Harvard University[8] and now boasts 500 million users, 70% of them outside the United States[7]. Twitter, despite its inherent brevity, breaks news stories[4] and the word “tweet” has entered the lexicon of many Internet users[1]. Justin Bieber, an international pop sensation, got his start singing covers of other artist’s songs on the video sharing community YouTube[15].

The cost to the privacy and welfare of individuals, however, has been great. User privacy has been compromised by confusing and invasive practices[18] and continuously more potentially sensitive information is being shared by social networks with third parties. Users have been fired for what they have said on their social networks[16, 3, 6, 19] and blogs[23, 13]. Lawsuits and criminal cases have emerged from comments and blog postings online where authors believed they were within their first amendment rights[20]. More gravely, harassment on Facebook[9] and personal ads on Craigslist[10] have led to tragic outcomes.

Personal information about users can originate from many different sources, some of which are under the user’s control and some of which are not. Users may publish personal information directly, they may accidentally reveal personal information through the aggregate of other benign information already available on the Web, their service providers may share personal or behavioral information about them, and friends and acquaintances may publish information that is libelous or unduly damaging to a user’s reputation.

Users are in the best position to manage the information they disclose themselves, but they need tools to make managing their privacy intuitive and efficient. Users are often unaware of the ramifications of their privacy and information disclosure decisions and, even if they understand the implications, they do not want

to spend a lot of effort specifying access control policies. In our previous work, we argued that content-based access-control (CBAC) policies enable users to express their intentions easily and succinctly. Our user study confirmed that CBAC achieves these goals[11]. We also demonstrated that, once the user’s policy is expressed in terms of content, machine-learning techniques can automatically infer how the privacy policy should be applied to newly-created objects.

Data generated and published by other entities, such as a user’s friends and acquaintances, is not under user control and cannot be placed under user control for both social and technical reasons. Furthermore, any mechanism for mitigating the damage from such information must advance the interests of the party that implements the mechanism. Thus, for example, we cannot expect search engines to filter content on behalf of the target user, since the search engine’s business interests compel it to provide the best and most complete search results that it can. To solve this problem, we enumerate the different parties involved in the disclosure of incidental personal information and consider possible mechanisms they can implement that protect users while advancing their other goals.

2 Privacy in Web 2.0 Services

Web 2.0 services collect personal data of many types and through a variety of channels. Bruce Schneier has proposed a taxonomy of social network information[22], which we reproduce below:

Service data: Data you give to a social networking site in order to use it (e.g. your legal name, age).

Disclosed data: Data you post on your own pages: blog entries, photographs, and so on.

Entrusted data: Data you post on other people’s pages that you lose control over once you post it.

Incidental data: Data people post about you (e.g. a paragraph about you)

Behavioral data: Data the site collects about your habits via behavior on the service

Derived data: Data about you that is derived from all the other data.

This taxonomy illustrates three sources for dissemination of personal information on the Web 2.0: the user himself, others, or from inference. Therefore, privacy can be compromised from any one of these three sources.

Web 2.0 services have provided many different mechanisms for users to control the distribution of information they post themselves. Some services support only a public/private distinction. Others enable users to password protect individual posts. The most common form of access control is the “friends” model, which lets users designate individual information as visible to nobody, friends-only, or world-viewable. Recently, some services have started to let users organize their friends into groups and assign different policies to each group[17]. Services have also offered several different schemes for organizing content to make specifying privacy preferences easier. For example, services may group content by type (e.g. photos vs. blog posts) and force users to specify a single privacy policy for all the information within each group. This is easier than specifying a policy for each object, but it may force users to choose a policy that either exposes more information than they’d like or hides some information they’d like to publish.

Behavioral and incidental information have posed their own problems. Facebook caused an uproar when they introduced status streams that simply increased the visibility of information that was already public. More recently, Facebook has taken steps to enable users to manage incidental information. Users can limit access to photos of themselves, even if the photo was posted by someone else. Also, Facebook allows users to control which friends can view content posted on their wall (a section of a user profile to which others can write messages). The incidental information problem cannot be solved by just one service, though – users still have no recourse when content is posted about them on sites over which they have no control. Facebook has also attempted to help users limit inference about themselves by controlling their visibility in Google and Facebook search results.

Although Facebook has made great strides to protect individual’s privacy, users still experience privacy invasions. Users are often unable to see the consequences of their actions. For example, several employees have lost their jobs after criticizing their employer on their blog or social networking profile[16, 3, 6, 19].

Studies have shown that users do not understand or check for security mechanisms[21]. Therefore, it is quite possible that most users are not aware of the full ramifications of their actions.

Lastly, the sophistication and reach of search engines endanger individual’s privacy and online reputation. One particular instance that drew the attention of the news media concerned a female law student at Yale University[14]. She was the subject of derogatory comments on the AutoAdmit law school discussion board. The law student only learned of the online discussion through a friend and after her job search resulted in no offers, despite her having comparable credentials to other successful recent graduates. Although she could not conclusively prove that the comments had prevented her from being hired, it is quite likely according to the Ponemon Institute[14] which found that half of U.S. hiring officials use search engines to investigate potential employees, with one third of searches yielding information that will deny the applicant a job.

We suggest two different approaches to improving user privacy. First, we explore how services can better inform users of the consequences of their actions and provide users with easier-to-use and more flexible controls over the visibility of their content. Second, we suggest reactive tools so that users can quickly identify privacy incursions initiated from other peoples’ postings.

3 Privacy Controls for Disclosed Data

As mentioned previously, Web 2.0 service providers have already made numerous attempts to provide their users with privacy controls for specifying who can see what content, but the controls have been either to simple and inflexible or too complex and difficult to use. Designing a good privacy preferences system is challenging because such a tool must simultaneously achieve several goals. First, users should be able to specify viewers, even if some members of their audience are not users of the same service, since otherwise they may be forced to make information publicly visible to reach those viewers. Secondly, privacy controls should enable users to write succinct policies that apply to large content collections, since users may create online profiles with large amounts of diverse data. Third, these mechanisms should provide flexible access control policies, since different users will have very different privacy goals. Fourth, whenever possible, the tool should help safeguard the user by inferring the privacy policy on newly created content, since users may often forget to specify a policy for new items.

We have previously proposed content-based access controls to solve these problems. In a content-based access control system, privacy policies are expressed in terms of the salient features of content, and therefore can be applied to many individual data items automatically. Concretely, we propose that privacy tools in Web 2.0 services should enable users to express their privacy preferences in terms of tags they place on their content (e.g. only my “college buddies” can see posts marked “Stony Brook University”). Tag-based access controls have several advantages in the context of Web 2.0 applications. First, users already tag the data they post, so tag-based policies do not require users to maintain any new meta-data. Second, users can tag objects however they see fit, so tag-based privacy policies are extremely flexible. Third, tag-based policies are portable across services. Since almost every Web 2.0 service supports tags, tag-based policies are a good foundation for enabling users to move their profiles from one service to another without losing their privacy policies. Similarly, tag-based policies will make it easier to create “sticky” policies that stay with content as it moves from one web service to another.

We have also shown through user studies that tag-based policies are easy to use and that users can create succinct privacy policies using the tags they already have on their content[11]. We first analyzed blogs with private posts on the WordPress blogging platform. WordPress allows users to password protect posts to restrict their visibility and access, but WordPress does not hide the tags on restricted posts, enabling us to compute the optimal equivalent tag-based privacy policy for any WordPress blog. We found that over half the users could express their privacy preferences (in terms of either the post being public or private) with less than 5 tag rules, and 75% could express their privacy preferences with just 9 rules. These results demonstrate that users can succinctly express their privacy preferences in terms of the tags they are already using.

We implemented a CBAC system for blogs called Plog that allowed users to express their privacy pref-

erences using rules on tags. We also allowed the user to override the tag-based privacy policy for individual posts. We then performed a userstudy on 28 participants to determine the usability of the tag-based policy language. The basic goal of the user study was to have the subjects read a blog constructed from actual WordPress posts and apply realistic privacy preferences in order to restrict the visibility of personal posts. Subjects first read the blog and then were asked to implement several privacy preferences described to them in plain English. Subjects were trained to use both the tag-based privacy tool and the per-post privacy tool, and they were free to use either approach to implement the privacy tasks. Despite the novelty of the tag-based system, about a third of the subjects chose to use it exclusively for all the tasks. We found that subjects that used tag rules created policies that were just as accurate as manually specifying access per-post. Subjects utilizing tag rules applied privacy preferences twice as fast as those using the per-post mechanism. And most encouraging, despite unfamiliarity with the blog content, blogging process and over 100 tags to choose from, subjects wrote near-optimal tag-based policies with respect to the number of tag rules and exceptions.

Tag-based privacy policies also enable services to provide two new and desirable features simultaneously: automatic tagging and automatic privacy inference. When users create a new blog post, they apply an average of 10 tags, 9 of which are tags the blogger has already used. Thus, machine learning algorithms can do a good job of suggesting tags for a new blog post by gleaning tag suggestions from similar posts from the past. We built an auto-tagger that had precision and recall scores of over 60%, substantially out-performing previously-proposed auto-taggers. With tag-based privacy policies, auto-tagging functionality immediately yields automatic privacy policy inference. Thus, we can achieve nearly effortless privacy policy maintenance.

Automated privacy policy inference can ease the burden of policy maintenance, but no inference engine is perfect, so we must design the privacy user interface to encourage users to review the results of the inference engine. By building privacy policy inference on top of tag inference, we already engage users since they will want to review the suggested tags for accuracy. Furthermore, we can integrate the privacy inference results into the standard work flow for posting a new blog entry or other piece of personal data. The privacy tool can also help users consider the consequences of their actions by summarizing the policy for each new object in plain English, such as “This post will be visible to everyone in the world.”

4 Privacy Controls for Incidental Data

We must identify the participants and their roles to better understand how incidental privacy disclosure occurs. This type of privacy disclosure happens when a searcher comes across sensitive information about an individual that the individual would not want the searcher to access. The participants in incidental privacy disclosure are the subject of the sensitive information, the author of the content containing the sensitive information, the searcher, the content provider and the search engine. Although the search engine could be part of the site, we will assume that it’s objective to find content and present quality search results is independent of the content provider’s goal to collect and publish content (for example, Facebook is the content provider and Google is the search engine). We must note that the only participant that is truly invested in protecting the privacy of the subject is the the subject himself. The content author is solely interested in the dissemination of their content, regardless of the privacy of the subject. A malicious searcher will try to find anything salacious to use for their own devices. A neutral searcher may be amenable to helping preserve the subject’s privacy by discarding the information, but does not necessarily have the information or recourse to verify the factuality or intent of the content author or the content provider. The content provider’s main objective is to serve content, and will realistically only become in a privacy dispute if legally compelled to.

Therefore, the only participant likely to aide the subject in protecting his privacy is the search engine because it’s objectives are not inherently incompatible with the subject’s desire for privacy. The search engine’s main focus is to generate quality search results. This conception of quality can include constraints that simultaneously serve the interests of the subject and the searcher. For example, a potential employer is most interested in comments on a person’s professional activities, so a search engine that priorities pages with professional over personal information can simultaneously better serve the employer while also subtly

encouraging the employer to make his decision based only on professionally-relevant data.

How can search engines and individuals work together to prevent incidental leakage of personal information? We call tools to prevent the incidental privacy disclosures as reactive because they have to react to information that is released. There are two basic scenarios for the operation of reactive tools. First, the search engine alone can annotate and modify the results it presents to searchers. Second, the individual can express their privacy preferences to the search engine, which the search engine takes into account when returning search results.

The search engine itself can employ several strategies to minimize incidental privacy disclosures. Search engines currently have the capability to differentiate between named entities (i.e. persons, places, things). Therefore, if a searcher queries the search engine about an individual, it can invoke a different algorithm for constructing and presenting search results. First, search engines could annotate search results with a ranking that gauges the trustworthiness, authenticity and objectivity of web sites. This will hopefully allow searchers to make more informed choices on either visiting the page or disregarding the resulting information. Search engines can also discard dubious pages or reorder returned results to highlight more authoritative and genuine results. Search engines can also use well-known algorithms for estimating the sentiment of a page, i.e. whether it is generally positive or negative and the intensity of its opinions. Search engines can also analyze web page to recognize discussion boards and reduce the rank of those pages when a searcher submits a query for a named individual.

Search engines could also withhold results from a searcher unless the searcher demonstrates why he should be entitled to potentially sensitive results. First, the searcher could present attributes in the form of a credential to show that they have something in common with the individual. The search engine would allow the searcher to access only those search results related to the attribute. The searcher could also demonstrate some awareness of a relationship between the user and an entity, entitling him to the related search results. For example, if an individual is a member of a particular social club, to retrieve all information based on this relationship, the searcher must provide keywords including the name of the club and the individual's name. A HR official performing a Google search on the individual will not have access to this result because he will be unaware of the potential hire's affiliation with the club. Neither of these approaches may be perfect in preserving the privacy of an individual, but would require more effort on behalf the the searcher to find personal information.

Search engines could allow users to provide some input into the ranking of search results. Users could communicate their privacy preferences to search engines using standardized protocols so the search engine can verify the individual's identity. Individuals, assisted by software programs, could encode their privacy preferences into a machine readable format for search engines to consider when returning search results. We imagine that search engines will heed individual preferences when returning polarizing, unsubstantiated and questionable pages. Their input may not be as influential over more objective and reputable web pages.

The users could provide the search engine with either general or specific privacy preferences. General privacy preferences refer to attributes and topics that could be associated with anyone. For example, an individual may express to a search engine that they would not want to share their location, religious and political preferences with anonymous searchers. In terms of specific privacy preferences, these relate to topics (represented by keywords) that are associated specifically with the individual. For example, if an individual has a profile with service he does not wish to make public, he can express to the search engine to not associate the service or any content on his profile with searches about him.

References

- [1] About.com. What is a tweet? <http://webtrends.about.com/od/glossary/g/what-is-a-tweet.htm>.
- [2] Amazon. Amazon mechanical turk. <https://www.mturk.com/mturk/welcome>.
- [3] Applicant.com. How to lose a job via facebook in 140 characters or less. <http://applicant.com/how-to-lose-a-job-via-facebook-in-140-characaters-or-less/>.

- [4] Claudine Beaumont. New york plane crash: Twitter breaks the news, again. <http://www.telegraph.co.uk/technology/twitter/4269765/New-York-plane-crash-Twitter-breaks-the-news-again.html>.
- [5] Justin Becker and Hao Chen. Measuring privacy risk in online social networks. In *W2SP 2009*, 2009.
- [6] ESPN. Facebook post gets worker fired. <http://sports.espn.go.com/nfl/news/story?id=3965039>.
- [7] Facebook. Statistics. <http://www.facebook.com/press/info.php?statistics#!/press/info.php?statistics>.
- [8] Facebook. Timeline. <http://www.facebook.com/press/info.php?statistics#!/press/info.php?timeline>.
- [9] Russell Goldman. Teens indicted after allegedly taunting girl who hanged herself. <http://abcnews.go.com/Technology/TheLaw/teens-charged-bullying-mass-girl-kill/story?id=10231357>.
- [10] Abby Goodnough and Anahad OConnor. Details released about 'craigslist' suspect. <http://www.nytimes.com/2009/04/22/us/22boston.html>.
- [11] Michael Hart, Claude Castille, Rob Johnson, and Amanda Stent. Usable privacy controls for blogs. In *Computational Science and Engineering, 2009. CSE '09. International Conference on*, volume 4, pages 401–408, aug. 2009.
- [12] Michael Hart, Rob Johnson, and Amanda Stent. More content - less control: Access control in the web 2.0. In *Web 2.0 Security and Privacy Workshop (W2SP'07)*, 2007.
- [13] Curt Hopkins. Statistics on fired bloggers. <http://morphemetales.blogspot.com/2006/10/statistics-on-fired-bloggers.html>, October 2006.
- [14] Ellen Nakashima. Harsh words die hard on the web. http://www.washingtonpost.com/wp-dyn/content/article/2007/03/06/AR2007030602705_pf.html.
- [15] ABC News. Pop star justin beiber is on the brink of superstardom. <http://abcnews.go.com/GMA/Weekend/teen-pop-star-justin-bieber-discovered-youtube/story?id=9068403>.
- [16] Charlotte Observer. Kirsten valle. <http://www.charlotteobserver.com/2008/11/16/354729/bosses-checking-up-on-workers.html>.
- [17] Nick O'Neill. 10 privacy settings every facebook user should know. "<http://www.allfacebook.com/facebook-privacy-2009-02>.
- [18] Barbara Ortutay. Facebook to end beacon tracking tool in settlement. http://www.usatoday.com/tech/hotsites/2009-09-21-facebook-beacon_N.htm.
- [19] Erik Palm. Facebooking while out sick gets employee fired. http://news.cnet.com/8301-1023_3-10228434-93.html.
- [20] David G. Savage. Blogger beware: Postings can lead to lawsuits. <http://www.latimes.com/news/nationworld/nation/la-na-blogger-suits-20100823,0,5604043.story?track=rss>.
- [21] Stuart E. Schechter, Rachna Dhamija, Andy Ozment, and Ian Fischer. The emperor's new security indicators. In *SP '07: Proceedings of the 2007 IEEE Symposium on Security and Privacy*, pages 51–65, Washington, DC, USA, 2007. IEEE Computer Society.
- [22] B. Schneier. A taxonomy of social networking data. *Security Privacy, IEEE*, 8(4):88–88, jul. 2010.

- [23] E. Simonetti. I was fired for blogging. Available from http://news.cnet.com/I+was+fired+for+blogging/2010-1030_3-5490836.html, December 2004.
- [24] Mark Snyder. Self monitoring of expressive behavior. *Journal of Personality and Social Psychology*, 1974.
- [25] N. Vyas, A.C. Squicciarini, Chih-Cheng Chang, and Danfeng Yao. Towards automatic privacy management in web 2.0 with semantic analysis on annotations. In *Collaborative Computing: Networking, Applications and Worksharing, 2009. CollaborateCom 2009. 5th International Conference on*, pages 1–10, nov. 2009.