Privacy Threat Model for Data Portability in Social Network Applications

Stefan Weiss
Johann Wolfgang Goethe-University, stefan.weiss@m-chair.net

Follow this and additional works at: http://aisel.aisnet.org/amcis2008

Recommended Citation
Privacy Threat Model for Data Portability in Social Network Applications

Stefan Weiss
Johann Wolfgang Goethe-University
stefan.weiss@m-chair.net

ABSTRACT
The advent of the participatory Web and social network applications has changed our communication behavior and the way we express ourselves on the Web. Social network application providers benefit from the increasing amount of personally identifiable information willingly displayed on their sites but, at the same time, risks of data misuse threaten the information privacy of individual users as well as the providers’ business model. From recent research, this paper reports the major requirements for developing privacy-preserving social network applications and proposes a privacy threat model that can be used to enhance the information privacy in data or social network portability initiatives by determining the issues at stake related to the processing of personally identifiable information.

Keywords
Privacy, Social Network Applications, Data Portability, Social Network Portability.

INTRODUCTION
Privacy has been discussed in various forms and settings for more than one hundred years by lawyers, philosophers, sociologist, psychologists, economists, technicians, politicians and other stakeholders. The Warren and Brandeis Harvard Law Review opinion piece in 1890, defining the right to privacy as “the right to be let alone” has set the direction for most privacy laws existing today. Alan Westin’s definition for privacy in 1967 “being the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others” added individual self-determination to the equation and constitutes the basis for privacy legislation such as the EU Directive1. Forty years later, with the advent of the participatory web and social network applications, individual self-determination once again seems to be the appropriate choice for dealing with privacy-enhanced web applications.

The extensive display of personally identifiable information (PII2) by users of social network applications (SNAs) on the Internet has raised concerns for privacy advocates. Additionally, while an increasing amount of privacy abuses via social network applications such as unwanted exposure, distortion, badmouthing, identity theft, cyberbullying or reputational damage become known, the demand for serious controls to protect the individual user from any damage increasingly comes from users themselves. At the same time, SNAs have been built on top of a technology that has not been set up with lots of inherent controls. The Internet’s original purpose was the openness of communication among a group of trusted people where privacy concerns did not exist.

This paper is based on preliminary research results from expert panel surveys being conducted currently among privacy and social network application experts. The study applies a Delphi survey technique based on structural surveys and makes use of the subjective-intuitive character of the participants’ answers. It is directed towards a carefully selected group of twenty-three technology experts who understand in depth the requirements and technical solutions for each of their specific areas of expertise, that is privacy law, security and privacy-enhancing technology, and the development of social network


2 Personally identifiable information (PII) is defined in this paper as being any information (a) that identifies or can be used to identify, contact, or locate the person to whom such information pertains, (b) from which identification or contact information of an individual person can be derived, or (c) that allows linking particular personal characteristics or preferences to an identifiable person.
applications. The surveys have explorative, predictive, and normative elements and additional survey rounds with the same expert panel provides the option to give feedback from results of the prior survey rounds and leaves room for validating answers and selected solutions.

Additional chapters point out the technical complexity of SNAs and the concept of data and social network portability is introduced and put in context to associated privacy risks. Finally, a privacy threat model for data portability in social network applications is proposed. It can be used to enhance the information privacy in data or social network portability initiatives by determining the issues at stake related to the processing of personally identifiable information.

REQUIREMENTS FOR PRIVACY-PRESERVING SOCIAL NETWORK APPLICATIONS

The rising use of SNAs on the Internet is a phenomenon that left lots of privacy advocates puzzled. Personally identifiable data that was sought to be at the core of any privacy-enhancing technology and needed to be encrypted, hidden or anonymized are nowadays provided willingly by users of social network sites (Kolbisch and Maurer 2006). Recent privacy studies for online communities, however, reveal the fact that most users are unaware of specific risks of privacy-invasive activities and have no idea to what degree their online profiles and the PII connected to it is visible and exposed to others (Acquisti and Gross 2006). A fact that also explains results from user surveys where users always say they are clearly concerned about their own privacy but then make decisions to reveal PII data about themselves that are contradictory to their concerns for privacy (Flinn and Lumsden 2005). Acquisti and Grossklags (2004) have elaborated on this dichotomy between privacy attitude and behaviour and concluded that individuals are neither able to calculate the probabilities and amounts of risks nor are they able to perceive the long-term risks and losses while acting in privacy-sensitive situations.

The usage of SNAs presents such a privacy-sensitive situation in which a great amount of PII is revealed to others. In our current research we presumed the social network users are not aware of the risks to their information privacy and, therefore, did not survey users’ preferences, attitudes and behaviour in more detail. Instead, we assembled a group of individuals who are considered experts in their respective fields of work and ask them about solutions.

Our research, based on a Delphi survey technique, had the objective to aggregate expert opinions and arguments on the most pressing challenges when trying to enhance the information privacy for users of social network applications. Privacy experts with a legal, technical and business background from countries in Europe, North America and Asia were asked to be on the expert panel. All of them have a particular experience in dealing with SNAs either from their academic, private company or public sector positions.

The research results are twofold. In a first expert panel survey with 41 structured questions, the major privacy concerns, the effectiveness of possible solutions, and the requirements for developing privacy-preserving SNAs among other topics were explored. A second expert panel survey will be conducted within the next two months and will validate proposed privacy solutions. From the preliminary analysis of the first expert panel survey the following factors depicted in Table 1 seem to influence the development and operation of privacy-preserving SNAs.

<table>
<thead>
<tr>
<th>Rank</th>
<th>Major Concerns</th>
<th>Possible Solutions</th>
<th>Requirements</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td>Having no control over usage and proliferation of PII</td>
<td>Complete transparency over the usage of one’s own PII</td>
<td>Privacy-by-design practices for web designers and developers</td>
</tr>
<tr>
<td>2</td>
<td>No transparency on what happens with PII</td>
<td>Privacy policies with an automated compliance assurance function</td>
<td>Transparent and open privacy handling practices</td>
</tr>
<tr>
<td>3</td>
<td>Unauthorized third party use of PII</td>
<td>Proactive and automated communication techniques on risks</td>
<td>Options for the user to easily report privacy invasions</td>
</tr>
</tbody>
</table>

Table 1. Factors influencing the development of privacy-preserving social network applications

Major Concerns

The experts were asked to pick the major privacy concerns they see in using SNAs. The top ranked answers for the major concerns are related to the lack of self-control and transparency to the user. These concerns support earlier findings where privacy is affected by the users’ inability to control impressions and manage social contexts, for example with the early introductions of such features like "News Feed" and “Beacon” in the Facebook application (boyd and Ellison 2007).
Furthermore, the surveyed experts see a major concern for the information privacy of users in the combination of an immature technology on the one side and providers on the other who need to proof their business model by further expanding ways to exploit the value of their users’ PII.

Possible Solutions

When asked about effective solutions it is interesting to note that proposed privacy-preserving solutions for using the Internet at large was quite different than the proposed solutions for SNAs. In the first case, traditional privacy-enhancing technologies (PETs) such as providing options to use pseudonyms, anonymization techniques, and data access rights management were priority. For social network applications possible solutions focused on transparency, automated compliance functions, and proactive communication techniques that can build awareness about potential risks. Those choices suggest that the expert group sees a distinct difference between Internet communication where traditional PETs can assure some level of privacy and social networks where the proliferation of PII is at the core of the underlying business model.

Requirements

The survey answers we analyzed for this paper are related to an importance ranking on a 5-point Likert scale. The experts were asked to rank the importance of a list of requirements for fostering the privacy-preserving use of social network applications. The result revealed that nearly 87% of the respondents see privacy-by-design practices for web designers and developers as the most important requirement for privacy solutions to work. This result is also supported by requirements set for ubiquitous computing systems where a comprehensive set of guidelines for designing privacy-aware ubiquitous systems were suggested (Langheinrich 2001). Further requirements for fostering a privacy-preserving SNA were transparent and open privacy handling practices and options for the user to easily report privacy invasions.

Other frameworks for analyzing requirements for privacy-preserving social networks have been suggested (Preibusch, Hoser, Gürses, and Berendt 2007). However, they concentrate on privacy in the sense of data protection, i.e. as a restriction on data access and data processing and not so much on the transparency and control mechanisms that need to be developed.

The preliminary findings of our first expert panel survey reveal an interesting point: the survey participants see the open nature of SNAs and their underlying database infrastructure as a given and suggest new forms of privacy-preserving mechanisms to solve the information privacy concerns inherent in such an environment. Privacy-by-design at the application development stage becomes more important. But the general call for more transparency, more structure and more control for the individual user is probably the greatest challenge for developers and providers alike. Especially because more than half of the surveyed experts attested an extremely low confidence level for the ability of providers to technically control all PII of their customers with current tools and procedures.

ADDING COMPLEXITY THROUGH DATA PORTABILITY

An additional dimension that represents potential risks to information privacy is added by the growing practices to make data from SNAs portable to other applications. The cumbersome activity of signing up for a number of new social networking services and the repeated work to enter profile information and to add friends to these sites has led to somewhat of a social network fatigue problem. Calls for more integration of applications and for a network of social networks stem in part from the inconvenience heavy social network users have when entering the same information again and again. In addition, SNA providers had to come up with a new business model in which they make use of combining data and deriving value from it.

The downside of data portability initiatives is the increasing complexity and disability to control data processes that involve PII. The initiatives such as OpenSocial or the Data Portability Workgroup have privacy on their agenda but they focus on providing simple privacy setting features that are not sufficient to the potential privacy risks.

Facebook’s chief privacy officer, Chris Kelly, said in an interview with the IDG news service (2008) that the problem with data portability comes in because there are all sorts of privacy and security worries [related to it], and there are a whole bunch of people out there who would gladly attempt to exploit somebody else's personal information if they could get one point of entry into a network, for instance, and try to export as much data as possible. And even with the own Facebook system that is connected to many third party applications, serious privacy risks have been discovered. A recent analysis of the top 150 Facebook applications by the University of Virginia, for example, has confirmed that 90% of these applications get access to private user data that they should not have access to (Felt and Evans 2007).

Our expert research confirms these worries with clear statements by the surveyed experts. When asked which technology developments they would see as the greatest challenges for the information privacy of Internet users, the ‘mashing up’ of applications and the underlying data combined with immature and often vulnerable technology was the number one concern...
with 65% of the survey participants. These developments coupled with the lack of accountable data management systems and auditable processes were on top of the discussion forum.

Even without the recent social network data portability initiatives, information on social networks can already be used through data mining and screen scraping applications that automatically infer real-world connections and discover, label, and characterize communities and individuals (Adamic and Adar 2003). Yet, Adamic and Adar stress the fact that the more data becomes portable and is exported to another application environment, the more likely information side effects occur. They describe these information side effects as by-products of data intended for one use which can be mined for another use in order to understand some tangential, and possibly larger scale phenomena.

An additional problem for preserving the information privacy of SNA users is the fact that most of the mash-up applications added to the SNAs are developed with open source software and, thus, also without clear privacy-by-design practices and defined privacy requirements. How to fulfill the call for more transparency, structure, and control in an application environment where no standards for privacy configurations have been set and where it is not clearly determined who has access to what data is an open question.

Google announced that it would adopt standards such as friend-of-a-friend (FOAF) and XHTML Friends Network (XFN) in their OpenSocial initiative to give developers access to the coveted social network graphs (the map of connections between friends) and related data (Techcrunch 2008). This allows developers and third parties to use PII as they please and integrate into their applications. At least, Google started the process of basing their data portability work on known standards. Others, such as Facebook, try to solve the issue simply by transferring all responsibility to the users themselves. Facebook’s current privacy policy3 points to contractual agreements with application developers in which they have to commit to respecting the users’ privacy settings but at the same time they state that they do not guarantee that all developers will abide by such agreements.

In order to understand the meaning of information privacy in the described context of making data and particularly PII portable in SNAs, a privacy threat model is proposed that draws on the following assumptions:

- Information privacy needs to be controlled on the data (PII) level
- The user needs to be able to determine the sensitivity and context of the PII provided
- Privacy-preserving data portability can only work if the user can earmark the PII provided with individual privacy preferences

![Figure 1. Setting Privacy Preferences](image)

Figure 1 describes the setting of privacy preferences by the user for each type of PII. Straight line arrows depict the control of the user for setting a preference by earmarking the PII in the application with a particular attribute. Dotted line arrows in the contrary show that there is no clear control by the user or not even by the application provider. Examples for privacy preference attributes the user should be able to set could be the following:

- a level of sensitivity of the PII for the user (e.g. the ‘marital status’ might be very sensitive for some users whereas it might not be sensitive at all for some users who are looking for a date with the ‘marital status’),

---

• a context in which the PII is supposed to be used (e.g. use of the association to the group ‘pop music lovers’ only in the context of providing signed-up private services such as a concert ticket alert service but not in a business networking context),

• a specific purpose for which the PII should be processed (e.g. collect the date of birth only for identification purposes but do not display it in the SNA profile unless otherwise authorized),

• an expiration date for certain types of PII (e.g. reconfirm the connection to a certain level of friends or contacts after a pre-specified time period), or

• settings for access and viewing rights of PII for specific individuals or groups (e.g. allowing all members of the user’s SNA network to view ‘marital status’ and ‘date of birth’ but not the association with the group of ‘pop music lovers’).

In the example in figure 1, the portability of the user’s data is enabled between application X and Y. Some data processing is taken place between application X and some mash-up applications. Some of those data exchange relationships are controlled (depicted as a straight line arrow), others are not under the application provider’s control (depicted as a dotted line arrow). But even if application provider X has all data flows under his control, the moment the data is exported to application Y, new data flows are taking place for which no control systems may exist. And even though application Z does not seem to have enabled data portability with other applications in this example, some third party mash-up applications may be able to access PII from the user through data mining or scraping techniques.

In order to stress the potential risks that are inherent in such a data portability case, a privacy threat model is proposed that attempts to visualize the most important privacy requirements to be build into any data portability project.

**PRIVACY THREAT MODEL FOR SOCIAL NETWORK PORTABILITY**

Potential threats to an individual’s privacy exist whether that individual provides PII to someone or not. Even if the individual would live as a hermit without an Internet and phone connection, there are potential threats to this person’s privacy if, for example, other people would start invading the hermit’s space for whatever reason. In order to preserve the hermit’s privacy, it is important to understand what constitutes a privacy invasion for the hermit (his own privacy preference), how likely it is to stumble on his piece of land without even recognizing it (public accessibility), and what degree of self-defense he has built up for himself (self-control). Of course, this example is quite trivial compared with the case of social network portability but it provides us with the basics for evaluating threats.

In order for protection mechanisms to effectively work in social network applications, all involved parties, and at the foremost the application developers, need to understand the potential threats that exist. Those could not only be threats to the individual user and the related PII but also to the business model of the social network application provider. The variety and seriousness of privacy threats in using social network sites have been pointed out in academic contributions and have especially addressed the inherent openness of social network applications (Weiss 2007).

The following multiple dimensions have to be taken in consideration:

**Individual Privacy Preference**

The most obvious dimension that comes to mind when dealing with information privacy is the setting of privacy preferences. Privacy preferences are always directly and subjectively linked to the user and are tagged as privacy preference attributes (e.g. a,b,c, etc. in Figure 1). They depend on a number of factors that create a certain level of concern for the individual providing his/her PII while using a particular technology in a specified context. The personal disposition of an individual towards privacy can depend on the person’s understanding of the technology used, the social background, the sensitivity of the data provided, past experience and as well as socio-psychological factors. Most SNAs provide some type of functionality that allows the user to set privacy preferences. Typically, it is the function to allow someone or a group of people access to the social network profile. However, SNA users should be provided with more than the profile access and viewing rights. Additionally, they should be able to set the context in which the data is supposed to be used, the specific purpose for which the data should be processed, and the expiration date the data might need to have.

The focus needs to be on each PII provided. Allowing access to the whole social network profile already falls short of the privacy preferences people have. I may allow one friend to see my profile but if a friend of that friend gets connected, I may not want that second or third degree friend to also view my profile. The more sensitive or confidential I regard a certain data set in a specified context the higher is the potential threat to my information privacy.
Ease of public accessibility

The ease of public accessibility is a dimension that most users of SNAs would not name as a fundamental factor in assuring their own information privacy because as we have pointed out earlier, the user awareness on the public accessibility or visibility of PII is very low. And yet, it needs to be a major responsibility of SNA providers to raise the awareness of users and making PII flows more transparent.

It could be argued that PII is already out there and, therefore, a social graph or data portability function only assembles what is already there. However, information privacy is about self-determination. The user who publishes PII to a Facebook or Xing site does not think of it as being public. Wang and Kobsa (2008) propose a solution on how to deal with public accessibility of PII on social network sites. The proposal includes a more transparent grouping of access rights according to pre-specified context. Here, with more transparency provided over the data being exposed in which context, the user’s awareness on his own privacy is also raised automatically.

Degree of user control over PII

The third dimension is the degree of user control over PII and is a fundamental part of the major privacy laws in Europe. It needs to be assured that the user can exercise control over each and every piece of PII – a difficult prerequisite if data portability means that data might get widespread without any means to control it. The contractual agreements laid out, for example, in Facebook’s privacy policy mentioned earlier are a clear sign for a very low degree of self-control.

Figure 2 describes the proposed privacy threat model for social network portability.

Figure 2. Privacy Threat Model

In order to design effective privacy-preserving mechanisms into a SNA that is based on data portability principles, each data set needs to be evaluated on the three dimensions of the privacy threat model.

(A) If the user would set the own privacy preference for example for a specific photo and tag it as “confidential” (the highest privacy preference level in the model), specific protection measures need to ensure that the photo is kept confidential. Possibly the user wants it to appear only for a brief period of time and only to one or a few individuals (for example for a diagnosis by a medical doctor). It also needs to be assured that the photo is not accessed by other applications and is not exported with the social network profile to other social network applications.

(B) The privacy threat would be very high if the photo tagged as ‘confidential’ would somehow be accessible to the general public, for example, because the user was not aware that everyone can see it if the photo is uploaded to the general profile photo stock. The protection measure that needs to be implemented in this case is an automatic warning to the user who is
about to upload the photo, providing the user with full transparency on where the photo is uploaded to and who can access it (also for example how searchable and visible it might be on the Internet). The user might believe that the data upload happens in a private and confined area when in reality it is uploaded in a general photo stock database that everyone can search with a specific mash-up application via the photo’s tags that the user entered.

(C) And finally, if the user has no control whatsoever over the uploaded photo because the application does not foresee any control mechanisms on the data level, the threat to the user’s privacy is also very high. The user has set his privacy preferences on “confidential” for the photo but because the application provider has various third parties who manage the photo stock and upload function and because he has not assured any technical means to enforce and transfer the individual’s privacy preference settings, the user’s information privacy is at stake.

FURTHER RESEARCH

The privacy threat model proposed here is only a start in supporting the design and development of privacy-preserving SNAs. The model can help in applying case studies to privacy-sensitive situations. We propose further research in building control and protection mechanisms that go beyond the current data protection and privacy view setting solutions for social network applications. As data portability initiatives advance, the management and control of privacy-sensitive data needs to be assured.

A range of technical concepts could be applied. Our expert research survey also asked the survey participants what kind of technical solutions or concepts should be invested in for developing privacy-enhancing solutions for social network applications. The following technical solutions were considered to be relevant for further research:

• Semantic technology (tagging data for context, purpose, and handling practices)
• Transparency-Enhancing Technologies
• Web Privacy Seals and Assurance Methods
• Using DRM techniques to develop a personal data rights management solution

Adopting semantic technology for marking PII as being privacy-sensitive and for applying privacy settings is probably the most effective way in the system environment of SNAs. In such solutions, it is important to develop systems that help automate certain user selections such as the privacy settings. Research related to the setting of privacy preferences using semantic schemata has been done by Berkovsky, Aroyo, Heckmann, Houben, Kröner, Kuflik, and Ricci (2007). In the presented approach, context-aware personalization is achieved by augmenting past experiences by the user with additional context-rich data. This might also be applied in deriving context-aware privacy preference rules and could help in reducing the input the user has to give in each transaction.

FUTURE CHALLENGES

It will be the main challenge in the future to develop privacy-preserving SNAs that will not kill the business model of social network providers. The PII of social network users is the most valuable asset for social network providers and it should also be in their interest to find solutions to protect those assets through effective means. Discussions on ‘who should own the data’ will most likely be replaced by ‘who can protect the data best’ and that is where users will engage in building up their social network profiles.

In addition, it can be expected that as more privacy invasions occur, lawmakers will also continue establishing consumer and privacy protection regulations that make providers of social network applications accountable for protection schemes. At least in Europe, recent developments by the EU and national regulators and courts have shown that privacy rights will not vanish but will be further strengthened as a basic human right especially in the realm of information systems. The German Constitutional Court, for example, just published a landmark ruling on 27 February 2008⁴, constituting a new “basic right to the confidentiality and integrity of information-technological systems” as part of the general personality and privacy rights in the German constitution. The ruling explains the relevance of using information-technological systems for the expression of personality and complements earlier landmark privacy rulings by the Constitutional Court that had introduced the “right to informational self-determination” (1983) and the right to the “absolute protection of the core area of the private conduct of life” (2004). The ruling defines the information-technical systems that are protected under the new basic right as being all

---

⁴ BVerfG, 1 BvR 370/07 vom 27.2.2008, Absatz-Nr. (1 - 333),
http://www.bverfg.de/entscheidungen/rs20080227_1bvr037007.html.
systems that "alone or in their technical interconnectedness can contain PII of the affected person in a scope and multiplicity such that access to the system makes it possible to get insight into relevant parts of the conduct of life of a person or even gather a meaningful picture of the personality" – a description that could be easily applied to social network profile data.

CONCLUSION

The proposed privacy threat model for data portability in SNAs summarizes the most important factors in developing privacy-preserving SNAs that include the portability of social graphs and PII among applications. Confirmed by results from a first round of expert surveys on information privacy in SNAs, the self-control of the user, more transparency over what is happening with processed PII and clear privacy-by-design best practices for application developers should be able to help in addressing increasing calls for protecting the information privacy of SNA users.

REFERENCES