The Need for a Paradigm Shift in Addressing Privacy Risks in Social Networking Applications

Stefan Weiss

Johann Wolfgang Goethe-University
Gräfstrasse 78, 60054 Frankfurt am Main/Germany
stefan.weiss@m-lehrstuhl.de,
WWW home page: http://www.m-lehrstuhl.de

Abstract. New developments on the Internet in the past years have brought up a number of online social networking applications within the so-called Web 2.0 world that experienced phenomenal growth and a tremendous attention in the public. Online social networking services build their business model on the myriad of sensitive personal data provided freely by their users, a fact that is increasingly getting the attention of privacy advocates. After explaining the economic meaning and importance of online social networks to eCommerce in general and reiterating the basic principles of Web 2.0 environments and their enterprise mechanisms in particular, this paper addresses the main informational privacy risks of Web 2.0 business models with a focus on online social networking sites. From literature review and current expert discussions, new privacy research questions are proposed for the future development of privacy-enhancing technologies used within Web 2.0 environments. The resulting paradigm shift needed in addressing privacy risks in social networking applications is likely to focus less on access protection, anonymity and unlinkability type of PET-solutions and more on privacy safeguarding measures that enable greater transparency and that directly attach context and purpose limitation to the personally identifiable data itself. The FIDIS/IFIP workshop discussion has resulted in the idea to combine existing privacy-enhancing technologies and protection methods with new safeguarding measures to accommodate the Web 2.0 dynamics and to enhance the informational privacy of Web 2.0 users.

1 Introduction

In the last few years, the Internet has seen new developments that not only changed the structure of some of the online business models as we know them but they will also change the way we see and use the World Wide Web in the future. Dale Dougherty coined the term Web 2.0 in 2004 and Tim O'Reilly[1] popularized the term later in 2005 as the "participatory Web" [1]. Compared to Web 1.0 (to apply the same terminology) when the Internet was used as a pure information source for consuming

[1] Both Dale Dougherty and Tim O'Reilly are leading the publishing firm O'Reilly Media Inc.

content, the Web 2.0 is now providing users with functionalities to actively participate and create content. Research and survey data [2-4] as well as anecdotal evidence in the form of newspaper articles or blog entries [5-8] see in these developments both opportunities and risks. This paper addresses the potential misuse of personal information in online social networking applications, referred to in this paper as informational privacy risk. After explaining the economic meaning and importance of online social networks to eCommerce in general and reiterating the basic principles of Web 2.0 environments and their enterprise mechanisms, important privacy research questions in online social networks are derived by aligning new privacy approaches specifically to the new dynamics of Web 2.0 applications. With the privacy research questions derived from the following discussion, this paper intends to raise awareness in enterprises and in the research community for the growing need to view and research privacy in the Web 2.0 environment differently than before and in developing new privacy-enhancing technologies to address informational privacy risks.

2 The economic value of online social networks

Online social networking websites such as MySpace, LinkedIn, Xing or Facebook typically provide applications for users to set up individual profiles, create virtual networks with friends and business partners, share articles, photos and videos, create content such as stories and blog entries, or to share opinions or preferences by giving online votes or setting search tags. Increasing online collaboration, interaction and personalization is the result – something that an online advertiser values as the source for more targeted marketing initiatives using sophisticated data mining capabilities.

Major acquisitions of social networking providers by investors in the past two years underpin the potential economic value of these firms. After News Corp. bought the social networking site MySpace for about half a billion US$ in 2005, Google acquired the video sharing site YouTube for 1.65 billion US$. Those acquiring firms see the commercial value of social networking sites like MySpace or YouTube not only in their attractive user base, the 18-30 year olds, but also in their potential influence on online retail growth overall. According to eMarketer Inc., online sales analysis data from last year's holiday shopping season in the U.S. for example supports the increasing commercial importance of social networks, blogs and user preference tags as word-of-mouth buying suggestions for small businesses [9]. Members of social networking sites become more active online buyers in response to preferences and "best of" lists displayed for example for music CDs within their community groups.

The online analyst company Hitwise underpins this trend by the growing percentage of online retail traffic coming directly from social networking sites – 6.2 % in the pre-holiday season in 2006 up from 2% in the same period in 2005. Hitwise sees in this data a clear proof that social networking sites such as Google's YouTube and News Corp.'s MySpace.com have begun displacing portals such as Yahoo Inc. as the new home base for Internet users. Social networking websites have emerged in the US market to become an integral part of web activity for many Internet users – in

September 2006, one in every 20 Internet visits went to one of the top 20 social networks, nearly double the share of visits compared to a year ago [10].

Analysts such as Forrester point out the attractiveness of users of social networking sites in more detail. In their report on "How Consumers Use Social Networks" from June 2007 [11], social networking site users come from households with an average household income of US$ 62,000 and above – quite an attractive consumer group. 50% of adult users and 67% of young users between the ages of 12 and 21 specifically state that they often tell their friends about products that interest them. Once all marketers have realized the potential of this user group and how to turn their online activities on social networking sites into their own benefit, it can be expected that the value of the users' profiles, their online behaviour and in turn the amount of all of their personally identifiable information will increase.

Attractive users have attractive personal data. As a result, the informational privacy especially for users of social networking sites is at risk. The following chapters will look at the challenge of assuring security and privacy for personal data on social networking sites and will also identify new research areas that can help to minimize these privacy risks in online social networks.

3 New privacy challenges and risks in Web 2.0

The increasing risk of misuse of personal data processed by online social networking applications is evident from computer science research [2-4] as well from anecdotal evidence in the form of newspaper articles or blog entries [5-8]. One example for the privacy risks users of Web 2.0 services see was expressed by a blogger named Jamais Cascio in October 2006 on the personal site Freds House which dedicates most of its blog topics to mobility, media and ubiquitous life topics. His blog entry reads as follows: "I'm feeling increasingly uneasy about my dependence on Google services. [...] I look around my desktop and I see Google Reader, Google Mail, Google Talk, Google Toolbar, Google Maps, Google Calendar, Google News, Google Analytics, Google Earth, and of course Google Google. [...] I think I need a new Google product to drop into beta. That would be, let's see, Google Data Privacy (GDP). GDP would allow me to review all of the information that Google retains on me across all services, from all devices, and from all sources. GDP would allow me to determine the maximum data retention period for each of my services. GDP would allow me to selectively opt out of cross-service data mining & correlation, even if it reduced the quality of the services I receive. GDP would allow me to correct any inaccurate data in my profile. And GDP would log and alert me when my data was queried by other services. [...] This is exactly the kind of thing that Google could do, should do, to maintain its "Don't Be Evil" motto, while compiling better -- more accurate and more useful -- information."

This blogger has described in length the main functionality that a privacy-enhancing solution in a Web 2.0 environment should provide, namely the self-control of one's personal data. It is clearly understood that more personal data collected, displayed, stored and processed in a decentralized environment and across multiple devices causes all sorts of concerns, one being the feeling to loose control. Risks

associated with this situation range from identity theft to online and physical stalking, from embarrassment to price discrimination and blackmailing [12]. The following table (Table 1.) lists a selection of privacy risks for the specific categories of social networking sites accumulated largely from published privacy breaches or from public discussions on the fears of such risks during the last 12 months.

Table 1. New Risks for Informational Privacy Emerge on Social Networking Sites

Category	Examples	Informational Privacy Risks
Business	LinkedIn Monster XING	Blackmail, Breach of Confidentiality, Data Reuse/Secondary Use, Discrimination, Aggregation (i.e. Pre-Screening for Recruiting, Harvard Business Case on Mimi Brewster)
Personal	MySpace Orkut Hi5 Classmates Bebo	Intrusion, Breach of Confidentiality, Data Reuse/Secondary Use, Aggregation, Identity theft, Abuse by Cyberbullies or Predators, Badmouthing, Pedophilia
Publication	YouTube Xanga Broadcaster Last.fm LiveJournal	Unwanted Exposure, Distortion, Data reuse/Secondary Use, Abuse by Cyberbullies or Predators, Video-bullying, Objectionable material, Pedophilia, Child pornography
Special Interests	BlackPlanet Cyworld Mixi WAYN Care2	Discrimination, Data reuse/Secondary Use, Aggregation, Intrusion, Exposure, Breach of Confidentiality
Individual	SecondLife Gaia Online	Exposure, Appropriation, Identity theft, Breach of Confidentiality, Insults, Cyberbullying

Considering the potentially differing interests of the data subject (here meaning the user providing personal data) and the receiving party in a commercial setting such as a social networking application, a definition of informational privacy that best describes the challenge to be solved is the following: “Privacy can be defined as an interaction, in which the information rights of different parties collide. The issue is of control over information flow by parties that have different preferences over information permeability.” [13] In this context, the individual user typically has particular socioeconomic motivations for a certain degree of privacy. According to Gary T. Marx, one of the leading privacy researchers in computer sciences, users may want to be protected from an unwanted intrusion of their time, space and person, they may want to see protection from discrimination or they may want to avoid “type casting” [14]. On the other hand, the provider of an online social network may have the interest to receive as much personal data as possible from an individual, including links to as many other people as possible, in order to increase the value of advertisement to his members. The more personalized the member profiles are, the more targeted and – in consequence – valuable adverts can be.

Looking back at traditional viewpoints on privacy protection in information and communication technology, technical privacy solutions tried to satisfy the socioeconomic privacy motivations of individuals predominantly through the use of privacy-enhancing technologies and identity management solutions [15]. Whereas those solutions address the user's anonymity, unlinkability, unobservability, or pseudonymity in form of a "protection and disguising" mode, these solutions may not address new privacy challenges a user faces when he openly and willingly displays a whole data set of personal information in form of his personal profile for example on a social networking website. In fact, hiding and disguising the personal data in the person's profile would most likely contradict with the purpose and perceived benefit of providing the personal information in the first place.

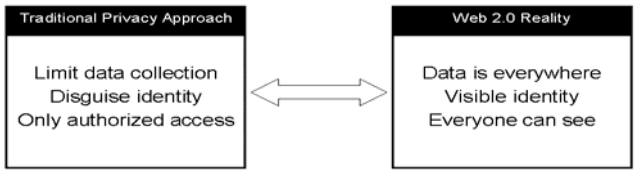

Fig. 1. Web 2.0 reality contradicts with the traditional privacy approach

Considering the general failure of the Web to satisfy requirements such as privacy protection, a balanced approach to intellectual property rights, and basic security and access control needs [16], additional privacy research in computer sciences will need to address solutions within the new "participatory Web". The Web 2.0 reality calls for a privacy paradigm shift adding privacy safeguarding measures for an open and decentralized environment. In this environment, the person whose data is at stake, may decide on a case-by-case basis if he wants to provide a certain set of personal information about himself in a specified context and if he only wants to provide it for a specific purpose and for a specific data receiver. In order to make those control features workable, the processes around those decisions and on what happens to the data need to be completely transparent.

Solutions for a policy-aware web such as the Platform for Privacy Preferences (P3P) or the Enterprise Privacy Authorization Language (EPAL) try to assure that personal data is being processed according to specified rules and policies. They offer the kind of tools that are needed to encode rules into web applications. On the other hand, they fall short on giving the actual control to the user, at least in their current application. If a user actual does set his privacy preferences using P3P, the system only checks against defined policies of the web site provider without any enforcement mechanisms. The actual control must be set down at the level of personal data.

Personal data is at the core of any online social network service's business model. That is why especially for this kind of application, privacy researchers need to go into more depth, looking at privacy safeguarding measures along the whole data processing life cycle, addressing the control and accountability of that data especially at the use end [17].

4 Privacy Research to Address the Web 2.0 Reality

Tim O'Reilly has defined the Web 2.0 as a "[…] platform, spanning all connected devices; Web 2.0 applications are those that make the most of the intrinsic advantages of that platform: delivering software as a continually-updated service that gets better the more people use it, consuming and remixing data from multiple sources, including individual users, while providing their own data and services in a form that allows remixing by others, creating network effects through an 'architecture of participation', and going beyond the page metaphor of Web 1.0 to deliver rich user experiences." [18]. In such an environment of decentralized systems and infrastructures that enable the quick and efficient development of systems, it is difficult to implement control features such as traditional security or privacy measures. Nevertheless, the rapid growth of Web 2.0 services is a reality and security and privacy research needs to adapt to it.

In order to derive relevant and specific privacy research questions in the new Web 2.0 environment, it is helpful to use the four principles and enterprise mechanisms of 'Wikinomics' [19], defined by Don Tapscott and Anthony D. Williams. There are a number of other more elaborate models and principles that could be used in the context of defining Web 2.0 dynamics, for example the "Web 2.0 Meme Map" [20] developed at a brainstorming session during a conference at O'Reilly Media. However, the author has purposefully chosen the Wikinomics principles here because they describe the relevant dynamics at work in the Web 2.0 somewhat more simplistically and they can easily be used to conceptualize the resulting privacy challenges and privacy research questions on a high level. While matching the principles of 'Wikinomics' and the respective privacy issues in this paper, the author has focused on the situation for an online social networking application and has not viewed different scenarios for example at video sharing sites or services that provide search and tagging functions. The case scenario of an online social networking service was identified earlier in this paper as being extremely vulnerable to privacy risks due to the nature of its business model dealing with personal data.

The principles of 'Wikinomics' are (1) Openness, (2) Peering, (3) Sharing and (4) Acting globally. Each of those principles motivate specific economic mechanisms within enterprises providing Web 2.0 services but each principle can also be related to specific privacy approaches discussed or recommended in current research papers as shown in the following table (Table 2.).

Table 2. Relating the principles of 'Wikinomics' and described enterprise mechanisms to privacy approaches

Principle	Enterprise Mechanism	Privacy Approach
Openness	Transparency	Accountability for data
Peering	Marketocracy	Informational self-determination
Sharing	Collaboration	Personal data property and usage rights
Acting globally	Multinational	Non-legal rules and policies

4.1 Evaluating each principle on its implication for the privacy of users of online social networks:

(1) Openness: If personal data is exchanged and processed openly in applications that are based on open standards and it is transparent who the involved parties are, privacy safeguarding measures need to assure accountability for the data and its authorized usage. It needs to be transparent to the user (transparency-enhancing technology) what happens to his data and it needs to be possible that each data process can be accounted for later on. The assumption here would be that data is being exchanged openly, thus, the requirement calls for a completely open process where the various parties can be made accountable for what they do with the data if necessary.

(2) Peering: The principle of "peering" builds on self-organization by a group of individuals. Applied to the case of an online social network service, individuals and groups of individuals determine the success or failure of the particular site by actively engaging for example in the linking of friends, building interest groups and communities and setting preferences that determine the exponential growth of the site. When thinking of the influence of the individual within a group and aspects of privacy, it is apparent that the individual needs to be provided with a function to determine what should happen with his personal data.

(3) Sharing: Sharing in the online social network setting means that the individual willingly wants to share data with others. That means for the service provider that he needs to provide collaborative tools to enable the sharing of data. However, when it comes to sharing sensitive personal data or providing data in a specific situation or context only, the individual might be reluctant to share with everyone and for any purpose. For this reason, privacy safeguarding measures need to attach something like a property or usage right to the personal data set. Lessons from digital rights management techniques or the concept of "sticky policies" for the Web might be useful to address this requirement.

(4) Acting globally: And finally the principle of "acting globally" brings up a range of issues when looking at privacy challenges in online social networks. Without legal boundaries of Web applications and even in some cases without any cultural boundaries and rules, it is a tremendous operational challenge that service providers face. How can rules for privacy aspects be set by each individual and how can they be enforced automatically? Legal and public policy regulations alone certainly cannot solve privacy challenges within those applications. Technology and privacy standards in the future may help to work on a common ground. Progress in the area of the web technology standards and the semantic web may also have some answers to privacy challenges in online social networks that are largely related to the specific context and usage.

The following table (Table 3.) attempts to give a brief overview of some of the privacy research questions that can be derived from the preceding discussion. The list of privacy research questions does not claim to be complete and, at this point in time, simply has the intention to raise awareness in enterprises and in the research community for the growing need to view and research informational privacy in the

Web 2.0 environment. In fact, it can be expected that interested readers, security and privacy experts can immediately add additional questions and topics to this list which should fulfil the underlying purpose of this paper to initiate discussions and thought processes around the topic.

Table 3. Inferring privacy research questions in the context of online social networks

Principle	Privacy Approach	Privacy Research Questions
Openness	Accountability of data use	• Is the definition and general perception of privacy in our networked world changing and how will privacy be defined in the future? • Do users see their privacy safeguarded if the data processes will be more transparent? • How do user groups and their behavioural patterns differ in open vs. closed online communities in relation to the type and extent of public display of their identity? • How can context-based data usage be integrated in existing Semantic Web concepts?
Peering	Privacy self-control	• How do group dynamics influence the attitude towards privacy? • Can we use existing literature on social network theory to explain aspects of trust and intimacy in online networking? • What is the commercial benefit of peer networks to eCommerce? • Would privacy self-control features in an online social networking site be perceived as a benefit and used as a solution to privacy concerns?
Sharing	Personal data property rights	• Under which circumstances and in what context are social network users willing to limit the usage of certain types of personal data (risk awareness)? • What kinds of gratification and cost models can show the value of sharing sensitive personal data with specific individuals or groups? • How can DRM technology be used by an individual for protecting his/her personal data from unauthorized access, copying, usage, or transfer?
Acting globally	Non-legal rules and policies	• What set of rules would users of online social networks see as essential to protect their privacy? • How can those personal, non-legal rules be converted into automated policies and attached to the personal data sets? (sticky policies concept) • Is it possible to derive general rule sets on privacy by studying different user groups attitudes toward privacy in different cultures and in different contexts or technology environments? • How can privacy standardization help to automate a privacy policy-aware Web?

4.2 Considering existing technologies for solving privacy issues in a Web 2.0 environment

Further research should evaluate and develop new solutions and methods that are able to ensure the informational privacy of individuals when using applications in the Web 2.0 environment. Privacy perceptions in the Web 2.0 have changed and will change further with the introduction of new information and communication methods. Besides researching those changing perceptions in terms of their social, psychological or economical roots, it should be of great value to look at existing privacy or security technologies and how they might contribute as a whole or in part to new privacy 2.0 solutions.

The following list of technologies or methods to be considered for evaluation against possible privacy 2.0 solutions (Table 4.) should be seen as work-in-progress. It served the audience of the FIDIS/IFIP workshop session on "Privacy and identity in social networks and online communities" as a source for discussion and could possibly be extended with ongoing work or planned work by the research community. The discussion during the workshop session led to the idea that combining existing privacy safeguarding measures with new methods to accommodate the Web 2.0 dynamics and bundling those into a packaged privacy 2.0 solution might have its greatest value by addressing an easier usability of privacy solutions at large, especially in an environment where users themselves increasingly participate.

Table 4. List of technologies and methods to be evaluated for their fit to solve privacy 2.0 issues

Privacy 2.0 Issues	Technology or method to consider for evaluation
Transparency and Accountability	• Audit trails and logs on data processes • Monitoring of pre-specified data usage • Privacy assurance methods (compliance) • Semantic techniques such as topic maps
User control model	• Trusted computing • Third-party service to manage personal data as a mediator • EPAL
Assuring the authorized usage of personal data	• Semantic web technologies adding usage context to personal data (tagging data) • Techniques from DRM solutions to be applied to personal data (Privacy Rights Management) • Watermarking techniques to mark the data owner (data provenance)
Managing privacy regulations and individual preferences	• Sticky policies concept (Semantic web) • Web site privacy with P3P • PRIME technology

The FIDIS/IFIP workshop session discussions have resulted in the viewpoint that besides the economic, social and legal questions around privacy protection in the Web

2.0 environment and particularly with online social networks, detailed technical research should be extended towards using semantic web languages, DRM technology and technology standardization to assure the informational privacy of individuals on the Web and the protection of personally identifiable information from misuse.

5 Conclusion

The growing economic value of online social networking sites in particular and Web 2.0 applications in general brings about new security and privacy risks that have not been adequately addressed by software developers, researchers and privacy advocates so far. Informational privacy risks such as identity theft, online or physical stalking, personal embarrassment, price discrimination or blackmailing differ widely among individuals and depend on the specific context. In the case of using online social networking services, the dominant approach to collect sensitive personal data at the outset makes it necessary to rethink traditional privacy approaches that were directed mainly at the protection and disguise of the user's identity information in the past. New privacy approaches need to direct their efforts to privacy safeguarding requirements that give control to the user. Data processes need to be transparent to the user, audit and monitoring methods need to be able to account for each data process and the pre-set privacy preferences of the user need to be managed and controlled diligently so that only authorized entities use the personal data for the specified purposes.

Research questions derived from the exercise of linking privacy approaches directly to the principles and enterprise mechanisms of Web 2.0 environments have shown that the pre-eminent goal for privacy research and PET development is likely to shift from access protection, anonymity and unlinkability type of solutions to privacy safeguarding measures that enable greater transparency and that directly attach context and purpose limitation to the personally identifiable data itself. Whereas specific research in this area needs to validate the need for new privacy approaches as described here, it can surely be concluded that the growth of online social networks and the systems that get developed around them need to get a stronger attention from the research community and from enterprises. It is clear that a number of new risks to information privacy arise where more personal data is collected, displayed, stored and processed in a decentralized environment and across multiple devices. More control, transparency and accountability can minimize those risks if all stakeholders put more attention on developing solutions in that direction.

References

1. O'Reilly, Tim, "What is Web 2.0", published on the O'Reilly website on September 30, 2005.
2. Nissenbaum, Helen, New York University, "Privacy as Contextual Integrity", Washington Law Review, v79 #1, Pages 119-158, 02-04-04.

3. Madden, Mary and Fox, Susannah, Pew Internet Project, "Riding the Waves of Web 2.0", October 5, 2006.
4. Cranor, Lorrie F., AT&T Labs-Research, "'I Didn't Buy it for Myself' – Privacy and Ecommerce Personalization", WPES'03, October 30, 2003, Washington DC, USA.
5. Heise Zeitschriften Verlag, Roth, Wolf-Dieter, „Tod im Netz: Wenn das Profil einer Social Networking Site zum Steckbrief wird", 09-19-06.
6. Wired News, Lynn, Regina, "The Internet makes us naked", March 9, 2007.
7. Time Magazine, Cox, Ana Marie, "Making mischief on the Web", 12-16-06.
8. Süddeutsche Zeitung Wissen, Stirn, Alexander, "Das soziale Netz: Ende der Privatsphäre", Ausgabe 13/2007.
9. eMarketer Inc., "Social networks influence online holiday shopping", Computerworld, December 25, 2006.
10. Hitwise Pty. Ltd., "Hitwise US Consumer Generated Media Report", November 2006.
11. Forrester Research Report, "How Consumers Use Social Networks", June 2007, Forrester Research Inc., Figure 6.
12. Gross, Ralph, Acquisti, Alessandro, H. John Heinz, III, Information revelation and privacy in online social networks, Proceedings of the 2005 ACM workshop on Privacy in the electronic society, November 07-07, 2005, Alexandria, VA, USA.
13. Noam, E.M., "Privacy and Self-Regulation: Markets for Electronic Privacy, in Privacy and Self-Regulation in the Information Age", 1997, US Department of Commerce.
14. Marx, Gary T., "What's in a Name? Some Reflections on the Sociology of Anonymity Title", Massachusetts Institute of Technology, 1999.
15. Hansen, Marit and Pfitzmann, Andreas: Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology, Version 0.28, May 29, 2006.
16. Weitzner, Daniel J., Hendler, Jim, Berners-Lee, Tim, and Connolly, Dan, "Creating a Policy-Aware Web: Discretionary, Rule-based Access for the World Wide Web" in Web and Information Security, Idea Group Inc., forthcoming.
17. Report in the 2006 TAMI/Portia Workshop on Privacy and Accountability, Massachusetts Institute of Technology, June 28-29, 2006.
18. O'Reilly, Tim, Founder and CEO of O'Reilly Media Inc., October 1, 2005.
19. Tapscott, Don, Wikinomics – How Mass Collaboration Changes Everything, Portfolio, Pages 20-30, December 2006.
20. O'Reilly, Tim, "What is Web 2.0? – Design Patterns and Business Models for the Next Generation of Software", Section 1. The Web As Platform, O'Reilly Media Inc., September 30, 2005.