THE IMPACT OF SOCIAL NETWORKS ON USER PRIVACY
What Social Networks Really Learn about their Users!

Steffen Ortmann and Peter Langendörfer
IHP, Im Technologiepark 25, D-15236 Frankfurt (Oder), Germany

Keywords: Social networks, Privacy, Web 2.0.

Abstract: Millions of users voluntarily release private and business data at community platforms without considering potential impacts on their real lives that may come along with that. Being used for personalized advertisement or profiling, user data are of utmost importance for economic success of the platform. Hence, platform providers exploit all promising options to gather data while privacy seems partially to be a pain for them. Beside data voluntarily released by the user, there are techniques and methods to secretly gather more user data, e.g., by proper fusion of miscellaneous information such as analysis of websites visited or social games played. In this article we investigate obvious as well as concealed data gathering options of platform providers. By that we uncover the true detailedness of user data collected by social networks to document our key message, i.e., social networks know EVERYTHING about their users. Finally, we discuss why existing privacy protecting solutions cannot stand up with the threats and risks resulting from easygoing use of social networks.

1 INTRODUCTION

Not only since Facebook^{TM}, online social networking is an ever-growing market where the big players already earn several million dollars per year. Online networks are part of the so-called “Web 2.0” providing ease of use for online information-, identity- and relation-management (Schmidt, 2006). Instead of pure information retrieval, the users generate and administrate online content on a platform provided. The platform mainly handles the necessary technical and organisational challenges. Social networks are a special type of those platforms supporting people with features for online interaction. Therefore the user establishes a mostly self-designed profile reflecting her/his online identity. Based on that, the user connects to other user profiles. Finally, the social network builds a graph, usually called “social graph”, where user profiles represent the nodes and links represent the edges respectively (Wasserman and Faust, 1994). The main functionalities of social networks are based on this graph, e.g., search engine and message services. According to a representative online survey, 63% of Internet users in Germany are already members of at least one social community (Rother, 2010). Within the group of 18 to 29 years old people, more than 90% are members of social communities. Nearly half of the users with profiles in Facebook, which in total maintains more than 500 million active profiles, login to the community almost every day. Anyway, communities still register an increasing number of users and community logins.

Private user data are of important value for community owners since they can be used (or sold respectively) for advertisement or even for better and highly profitable, personalised advertisement. The latter seems to provide market sizes of several billion dollars per year and hence, is of enormous interest for platform providers. Only a huge number of users and a maximum of personal user data assure attractiveness and economical success of the platform. Consequently, priority objective of platform providers is collecting as much user data as possible. Therefore platform providers exploit all promising options to gather data while privacy seems partially to be a pain for them.

Due to the rise of social networks large parts of human life and social interactions migrated into the Internet. Promoted by network providers and missing awareness for Internet-media, especially young users tend to reveal nearly their complete real life in a virtual community. Hence, a vast amount of private and business data has been put into and stored in these networks. Millions of users voluntarily release pri-

Ortmann S. and Langendörfer P..
THE IMPACT OF SOCIAL NETWORKS ON USER PRIVACY - What Social Networks Really Learn about their Users!.
DOI: 10.5220/0003402200210026
In *Proceedings of the 7th International Conference on Web Information Systems and Technologies* (WEBIST-2011), pages 21-26
ISBN: 978-989-8425-51-5

vate and business data without considering potential impacts on their real life that may come along with that. Platforms of course get aware of private data and social contacts explicitly given by the user. However, most users cannot even imagine how much private, social and technical data they produce by mere usage. Besides data explicitly given by the user, platforms are able to gather information that is not explicitly given but can be inferred from user behaviour or context. Even worse is the fact, that adverse information may also be published and linked by community contacts. There exist practical methods to kindly "force" users into betrayal of other community members they know. Thereby privacy of community members may be breached without being noticed by the affected user. However, we consider such behaviour contradicting to laws or privacy ethics at least.

Beside own experimental results, we have analysed the state-of-the-art and media reports concerning privacy issues, general terms and conditions as well as the official privacy policies of the big players in the market, e.g., of Facebook (Facebook, 2010a; Facebook, 2010b). In this article we list in detail the user data that are accessible by mere platform usage. We further focus on discovering methods secretly gathering more information about community members than those to which users explicitly acknowledged to be collected. By that we show the detailedness of user profiles, very often reflecting real life data, captured by the social network. Based on that we present existing privacy protecting technologies and discuss the reasons why these cannot bear up the privacy threats posed by social networks. Finally we conclude with identifying important aspects for further research and discuss potential provisions for privacy aware usage of social communities.

2 NO ROOM FOR PRIVACY?

Within the context of data gathering, people worried about privacy often are confronted with the slogan "I've got nothing to hide" (Solove, 2007). Even if this is true considering laws, it does not mean private data may be public in any case. Since privacy is your right (Warren and Brandeis, 1890) to self-determine when, how and to what extent information about yourself is communicated to others (Kuhlen, 1999; Westin, 1967), you should be in almost full control of your information. As in real life, private data must also be protected in online communities! Unfortunately, privacy protection mechanisms provided by the social networks feature significant drawbacks. These either provide too little options for securing any private data or are too complex with respect to usability or do not allow for fine-grained privacy settings. Pseudonymous usage of communities is also unpopular or not permitted by platform providers.

But privacy is more than just control of data. Good privacy practise starts with fair information about kind and purpose of data collected. The latter is the crux of the matter. While general terms and conditions provide more or less full descriptions of what data is collected, the purpose usually is not given. Likewise, informative value of data collected as well as inferences of those is difficult to understand for the standard user. Most users do not even approximately know the mass of information they provide. In the following we uncover the true detailedness of user data collected by social networks. Beside obviously gathered data voluntarily released by the user we thereby focus on the data gathered alongside usage as well as "methods" of concealed data acquisition. By that we document our simple (and scary) key message, i.e., social networks know *everything* about their users.

2.1 Who is the User?

Usually creating a user profile in a social community starts with giving basic data, such as name, birthdate, gender, language, address and photo. This is not much information but of course, this is enough data to determine where you live, to guess your social state because of your residential area, to search for your number in public telephone registers and to look at your house using free online street maps for example. Also uploading a user photo, which is highly recommended by all communities, allows to use face recognition systems, e.g., as freely available in Google's Picasa^{TM}, to search for other available data about the person depicted. If such technique is combined with mobile devices, everybody may obtain detailed data, in our case the user profile, about any unknown person just by taking a picture in public.

Data given by the user allows also getting a more individualised idea about the person behind the user profile. Hence, that data is even more valuable in view of advertising. Usually, the user may disclose status of relationship, political interests, hobbies and favourites. Of great value are user interests, hobbies and memberships in clubs. Marketing and selling companies invest lots of money to acquire knowledge about people's favourites in music, movies and books. The social network provider gets this information almost for free. Since such information is not always given within the user profile, there also exist community games asking for personal favourites to find other users in the community with same or similar interests.

Likewise community applications may ask to evaluate certain product or fashion. By that, the user is unconsciously quizzed about interests, opinions and personal taste.

2.2 Who does the User Know?

Data voluntarily released by the user is just the beginning of invasion of privacy. Contacts within the network best reflect a user. First, contacts are people that a user probably knows. Second, the platform exactly registers the contacts visited, the photos watched and messages sent to contacts. Here, exactly means to determine in detail how many messages a user writes to whom at what time. Even if usually prohibited by law, it is technically possible to access the content of the message, too.

Further, the platform registers how often and how long some activity has been carried out, e.g., how often a user visits which contact or how much time the user spends in reading articles or watching photos of certain user. Finally, the usage of "poke" options, a kind of digital greeting in Facebook, may reveal close relation- or partnership between users. By that, differentiating contacts between friends, family and acquaintances becomes possible. Obviously, there are simpler opportunities to gather this information because platforms also allow users to sort contacts into certain lists holding all friends, family members, etc.

2.3 What does the User Do?

Inference of comprehensive personalised profiles from data explicitly given by the user is less than the half of all data gathered. All platforms automatically generate very detailed usage analysis data by maintaining log files for each user. Log files contain product information about the browser used, IP address, date and time of access, amount of data transferred as well as referrer and target URLs. In other words, the platform registers how often and how long certain user accesses the platform. Determining referrer and target URLs enables also collection of information about external websites and services used. Logging of IP addresses and browser information allows identifying the accessing device or user. Since cookies are used to locally store encrypted credentials, it can be further determined, whether other users known to the platform frequently enter the network from the same device, too. Such pair or group of persons may be family members or live in any relationship to each other. Moreover, recorded data can be analysed and give semantics beyond pure technical information. Platforms exactly know what their users are doing online. They are also aware of how long and how often a user plays social games or uses certain offers. Sometimes the platforms cannot register what users are doing in third party offers. For that case, third parties usually have to confirm delivery of usage analyses at least. Platforms recognise the time a user is online, no matter whether it is in the morning, in the evening or at night. In general, everything the users do online is recognised by the platforms just as if somebody is directly observing the users. It does not matter who, where or when - the platforms always track what their users are doing.

2.4 Where is the User?

When a mobile device is used to access the network, the platforms are also permitted to identify the mobile device, the mobile service provider and yet actual position data, e.g., derived from an internal GPS module. This enables to track the user anywhere at anytime. Since it is a significant intrusion into privacy, automatic determination and publication of location data when using mobile devices is disputable. In addition, Facebook offers a relatively new service called Facebook Places. Facebook Places enables users to publish in real time where they actually are and hence, Facebook also gets to know the location data of users. This is of less concern if publication of location data is in the hand of the user. Unfortunately, this service allows publishing location data of other users, too. However, meanwhile Facebook has learnt about their user privacy requirements and allows prohibiting publication of location data by other users. Certainly, the user has to explicitly restrict such publications in the privacy settings, because these allow friends to publish location data by default.

2.5 What does the User Think?

Finally, social networks even know what their users think! They know results from polls and votes, log user-written comments and store twitter messages and blogs as well. They register the content and the kind of advertisements a user responds to. Last but not least, the platform providers get to know how concerned users are about their privacy. Providers learn whether a user apparently reads the general terms and conditions or just clicked the accept button. Of course, the platform provider exactly knows the privacy settings of a user as well as how often and for which purpose the user changes the settings. Thereby the provider learns consumer acceptance in detail and hence, can well estimate acceptance of platform enhancements or new data gathering methods.

2.6 What do other Users Think about the User?

During our research we found an interesting and likewise scary approach to gather more “user data”. Social networks usually provide third party offers, which predominantly are games or fun applications, called apps. Every user should be aware of the fact that apps are provided for no other reason than to make money or gather user data or both. Apps basically are free but allow collecting side data produced by their usage as well. Beside others, apps of mainly two types put a user’s privacy at risk, i.e., apps sharing personal interests between users as already mentioned and apps asking for data and opinions about other users. Especially the latter type of apps partially significantly invades into user privacy by sniffing out other users. As an example, we found the app “My friends 1.0” in the German social network StudiVZ^{TM}. This app asks users for personal opinions about contacts. While testing this app, we were asked questions such as “Do you believe Henry has enough sex?” or “Do you think Henry had been in jail anytime?” about a user named Henry (name substituted).

These are simple questions but obviously nobody would like respective answers be publicly available. The really treacherous issue with this app is that users are tricked into providing information about others by gaining a certain amount of chips when entering this data. This virtual currency can be used to buy answers their “friends” have provided about the user in the same app. So there is a kind of interest in getting chips to learn what others think about you. Unfortunately the app allows answering these questions even if the other person is mostly unknown. This was the case for Henry, who was a bogus person, but already made new friends and for those Henry already could answer those questions. Unfortunately, to the best of our knowledge there exists no approach that may protect users from being sniffed out in such a way.

2.7 What does the User do Outside the Network?

Despite the vast amount of user data collected within a social network already, the world’s number one social network Facebook opens up methods to gather user data from outside the network, too. Therefore Facebook established the Like Button. It enables Internet users to “like“ and thereby share web content by just clicking a button. Consequently, the inhibition threshold to share information that way is extremely low. However, the idea behind the Like Button is as great as scary. Integrating the Like Button, which is hosted at Facebook, into some website Facebook receives information about every entity accessing the website. This information contains the IP address, the previously visited website (referrer-URL), timestamp, browser id, etc. If the accessing person is logged onto Facebook at the same time, the Like Button displays other users who also liked this website and Facebook gets the identity of the accessing user, too. Anyway, Facebook receives all information mentioned whether or not the accessing entity is a Facebook user. Hence, there are still privacy risks since the accessing entity may be identified by existing cookies, the IP address or the browser used. In addition, the Facebook sharing analyse tool, which can be used for free by everybody, determines the number of likes of any website implementing the Like Button.

Obviously, Internet users could also share information and comments by email like they have done before Facebook, but using the Like Buttons is much easier and faster. The only difference is, that Facebook now also knows, which Internet content somebody consumed and what Internet users think about a website or any other content *inside and outside* the social network. With more than 500 million active users and roughly two million websites said to implement the Like Button, Facebook slowly but surely registers a large part of the worldwide Internet usage. As all roads lead to Rome in the real world, it obviously seems possible that all clicks in the Internet shall lead to Facebook in the future.

2.8 Collecting Data of Unregistered Persons

Meanwhile social networks have started collecting any data possible, even of persons who are not registered at the platform. Beside using the Like Buttons for such purposes, contact importer tools offered by the social networks are of prominent use. These tools access contact data available in Microsoft Outlook, on the iPhone or in MySpace to autonomously search for platform users with respective data. By that, these tools transfer any contact data like addresses and phone numbers to the platforms even though affected people are not registered users. Moreover, affected people are neither informed about usage of their contact information nor asked for permission to do so. Here the respect for someone’s privacy is completely at the responsibility of the user granting access to contact data. It further is at least questionable whether such application already contradicts the law. Even if this approach is yet hard to understand, the user partially has to provide the password required for accessing the email account. It seems almost incredible that

people not even realise they are offering access to all emails stored in the account, too.

3 NO TECHNICAL PRIVACY PROTECTION AVAILABLE?

The phenomenon of social networks clearly indicates that privacy enhancing technologies for Internet use as researched in the past are becoming useless. Encrypting data when communicating, or restricting user data revealed when using Internet services does not help if the data is voluntarily provided and accessible in social networks. Also techniques (Maaser et al., 2008; Cranor et al., 2006) based on P3P and APPEL, which allow to restrict information to be revealed to service providers or at least to inform service users about what data is collected and for which purpose, are not the correct means to deal with the privacy risks stemming from social networks. Even though fair information about kind and amount of data gathered would be a step into the right direction, our experiences clearly show that as long as social networks have no additional gain or legal restriction, they will restrict usage of privacy protecting technology to a very minimum.

Since this is work in progress we currently evaluate available as well as novel technical approaches for privacy aware usage of social networks. Technologies which might enhance privacy protection within social networks are distributed profile management as well as different strategies and tools enabling systems to configure several privacy profiles automatically. Distributed profile management as presented in (Langendörfer et al., 2004) may get rid of the necessity to have all user data centrally handled and stored in the social network. Instead, the communicating device used to access the network may negotiate fair information exchange with the platform. Likewise automatic assurance of individual privacy profiles of several users at the same time, as supposed for application in pervasive environments (Ortmann et al., 2008), may be a potential research domain. Obviously the most difficult part is to deal with the mass of users a social networks usually have. As a start, grouping users with similar privacy requirements may offer a chance to restrict exchange of data between different user groups as well as between groups and the social network according to the privacy regulations set for each group.

Of course, we are aware of the fact that even advanced privacy preserving technologies cannot prevent from disclosure of private data by other Internet users. In addition to researching novel technical solutions, here we clearly consider the social network provider and legislative organs to accept the responsibility of protecting privacy of Internet users. Therefore regulations forcing social network providers to adhere to privacy ethics and rules, e.g., as defined in (International Working Group on Data Protection in Telecommunications, 2008), must be declared. However, we strongly believe such privacy regulations to be achievable by means of law only.

4 CONCLUSIONS

In current ongoing work we have investigated the terms and conditions, the default privacy settings, as well as the type of offered apps and the data exchange more or less enforced by social networking platforms. Our research clearly shows that users of social networks are left alone when it comes to protecting their privacy. To provide evidence of our key message saying that social networks know *everything* about their users, we have presented the true detailedness of data gathered *inside and outside* of state of the art social networks. Currently available privacy protecting technologies, if applicable to social networks at all, clearly favour the platform providers, so the users are more or less tied to the rules dictated by providers. This is really unfair given the technical skills and awareness of the standard Internet user and the platform providers. However, what worries most are three facts:

First, the big players in the social network market, first of all Facebook, have started several approaches to gather data of unregistered Internet users as well, e.g., by hosting Like Buttons and providing contact importer tools.

Second, social platform providers do not clearly state privacy risks, especially about threats coming from additional means to collect and infer information from data voluntarily provided by the users.

Third, the amount of information, platform providers can access directly and indirectly, can only be guessed. Even though, we do not know for sure for which purpose the platform providers are using the user data.

The ultimate advice we can actually provide is, think twice before revealing data, when in doubt do not provide it. When asked for more data become suspicious and even more careful, and never ever reveal data of friends, relatives, etc. Given our research results we are sure that protecting privacy cannot be solely achieved by technical solutions. We consider

future privacy protection a multi disciplinary task. It requires expertise in social sciences, jurisdiction and computer science. We sketch what these disciplines can contribute to better privacy.

Social Sciences. The attitude towards privacy is currently changing. People are less concerned about providing their own data. Here social sciences can contribute by again rising awareness of what people might lose when providing data to others. To achieve this goal, potential and real consequences of sloppy privacy treatment need to be collected, summarised and reported.

Jurisdiction. The legal setting of terms and conditions should be analysed and new guidelines should be generated, which indicate a stricter handling of user data and are backing up the users' rights. In addition clear definitions are required on how sessions have to be protected. Finally, instead of mere guidelines for fair information practises, unified, standardised legal conditions and privacy rules need to be established across country borders.

Computer Science. Here two research directions seem to be very promising with respect to better privacy. First it would be a significant improvement if users become aware of what information can be inferred alongside usage or when revealing additional data. A tool which goes in that direction has been researched in (Ortmann et al., 2007). To generate the required information data mining tools can be adapted to provide the basic information. Then a tool for data deduction out of the provided and retrieved data needs to be developed. An open point here is whether platform users will be able or allowed to access all data required to do proper data retrieval and on top of that to deduce of potentially revealed data when providing additional data. Second decentralising social network platforms, as announced in the Diaspora project, would for sure reduce privacy risks since no single entity is in possession of all the user data. Consequently, information inference from user data is much more difficult. For such purpose, we have patented an approach enabling fair bilateral information exchange that does not favour certain entity. Privacy of users of Peer-2-Peer based services, as decentralised platforms would provide, can significantly gain from such an approach. The open technical question is whether or not such a distributed social network can provide the service, platform users expect. The other question is rather political or just business related. Would existing platforms allow such a new system to grow?

REFERENCES

Cranor, L. F., Guduru, P., and Arjula, M. (2006). User interfaces for privacy agents. *ACM Trans. Comput.-Hum. Interact.*, 13:135178.

Facebook (Retrieved September 30, 2010a). *Facebook Privacy Policy*.

Facebook (Retrieved September 30, 2010b). *Facebook Terms and Conditions*.

International Working Group on Data Protection in Telecommunications (2008). Bericht und empfehlung zum datenschutz in sozialen netzwerkdiensten - rom memorandum. Technical report, Berliner Beauftragter für Datenschutz und Informationsfreiheit.

Kuhlen, R. (1999). *Die Konsequenzen von Informationsassistenten*. Suhrkamp.

Langendörfer, P., Maass, H., and Falck, T. (2004). Plasmads: Smart mobiles meet intelligent environments. In *Proc. 4th Workshop on Applications and Services in Wireless Networks*. IEEE Press.

Maaser, M., Ortmann, S., and Langendörfer, P. (2008). *The Privacy Advocate: Assertion of Privacy by Personalised Contracts*. Books of selected papers from WEBIST conferences. LNBIP Springer.

Ortmann, S., Langendörfer, P., and Maaser, M. (2007). Enhancing privacy by applying information flow modelling in pervasive systems. In *OTM Workshops*, Vilamoura, Algarve, Portugal. International Workshop on Privacy in Pervasive Environments (PiPE07), Springer LNCS.

Ortmann, S., Langendörfer, P., and Maaser, M. (2008). Adapting pervasive systems to multi-user privacy requirements. *International Journal of Ad Hoc and Ubiquitous Computing*, Vol. 3(No. 1):8597.

Rother, P. (2010). *Web 2.0 Communities: Geschäftsmodellanalyse und Erfolgsfaktoren*. BoDBooks on Demand.

Schmidt, J. (2006). Social Software. Onlinegestütztes Informations-, Identitäts-und Beziehungsmanagement. *Forschungsjournal Neue Soziale Bewegungen*, 19(2):3747.

Solove, D. J. (2007). Ive Got Nothing to Hide and Other Misunderstandings of Privacy. *San Diego Law Review*, 44:745.

Warren, S. and Brandeis, L. (1890). The right to privacy. *Harvard Law Review*, pages 193220.

Wasserman, S. and Faust, K. (1994). *Social network analysis: Methods and applications*. Cambridge University Press.

Westin, A. F. (1967). *Privacy and freedom*. Atheneum, New York.