Abstract

Privacy and security features are definitely not the driving forces in the creation and mainstream adoption of online communities. Most people taking part in online social networking (OSN) appear to be unaware of the serious security and privacy implications of sharing their personal information and experiences online, and thus there is no real demand for such innovation. This ignorance and lack of proper security and privacy models opens the door to would-be attackers. Indeed, with their facilities for exchanging messages and sharing content with the other members of the community, they provide a platform for many forms of online crime. This paper gives an overview of the privacy and security risks involved in the social web and discusses some possible solutions.

KEYWORDS: social networking site, online social networking, privacy, security, social web

1 Introduction

Sites dedicated to OSN are among the most trafficked spots on the Internet. According to a report by comScore Media Metrix the popular social networking site (SNS) Facebook¹ was the 16th most trafficked web property in the US in January 2008 attracting over 33 million unique visitors [9]. Moreover, another report shows that online social networking is also fast growing in popularity: compared to June of 2006 by June of 2007 the average number of daily visitors to Facebook.com had increased by 299% and similar rapid growth was visible in the numbers of its competitors, such as the 72% increase of MySpace² to 28 million visitors [11].

OSN is also a global phenomenon, with different types of services competing for subscribers in a global expanding market [25]; some are more geared to providing business connections and others center on match making. For example, aSmallWorld³ is an invitation-only SNS intended to serve the needs of the social elite. The popularity of specific services varies greatly depending on the geographical location [5]. Some examples of highly popular sites are MySpace, Facebook, Orkut⁴, Bebo⁵, Friendster⁶ and Cyworld⁷.

²http://www.facebook.com/
³http://www.myspace.com/
⁴http://www.asmallworld.net/
⁵http://www.bebo.com/
⁶http://www.friendster.com/
⁷http://us.cyworld.com/

As with all communication technology, privacy and security are central issues. SNSs provide the users with an online presence and capabilities for various types of interaction with the other people taking part. The issue is that people tend to share a lot of information about themselves in this more or less public forum. There seems to be a disconnect between user’s perception of privacy and the framework that is actually in place. They might also be encouraged to share information as much as possible, to egoistically promote themselves. For example, a popular video sharing service YouTube has the tag line "Broadcast yourself" in its logo ⁸. The apparent control over their data and ease of sharing also contribute to the outcome. Users also believe that the benefits outweigh the potential harm. They do not understand that even small pieces of information when put together may be damaging, and that hiding behind pseudonyms is not a sufficient safety precaution: profiles can be identified and tied to a real person, e.g. by combining and comparing information from different sources.

What is especially important to grasp, are the differing characteristics of online and offline conversation. Once something appears on the Internet, it’s almost impossible to remove. Also, it is a cumulative information source with quick and wide distribution capabilities. After all, privacy is not always so much about secrecy but control, and in the Internet you have very little of that due to automated replication, caching, archiving, aggregation and indexing. An offline friend will use common sense when deciding whether or not to relay forward some information shared with him, an online connection generally will not, the information is automatically and invariably propagated through the network of friends and broadcast to all of them.

In contrast to their behavior online, people do seem to value privacy if anything is to be concluded, for example, from the uproar caused by Facebook’s news feed [34, 28]. This feature automatically aggregates any changes in the profiles of friends into a handy feed. The protest is actually more indicative of the predominant false sense of privacy. Users seem to think their profiles feature information they would not like to be actively distributed but still enable exactly that by storing it online.

The remainder of this paper is organized as follows. We first elaborate the significance of the issue. Next, we give an overview of the specific threats and discuss both their implications and proposed solutions. Finally, we summarize the matter, and conclude by listing some topic areas for future work.

⁸http://www.youtube.com/
2 Social Networking Phenomenon

The Web 2.0 technology ushered in an era of user-created content. The reinvented web provides a framework for user-driven applications with provisions for effortless user interaction and information sharing. At the heart of this information revolution that is the social web are the services known as social networking sites (SNS), which facilitate online social networking (OSN).

SNSs build on two basic concepts, the profile and the network of relations [16]. Profiles are personalized pages with information about the owner, shared with others on the network. Through their profiles, individuals express their tastes, interests, world views, opinions, and lifestyle by means of bulletin boards, blogs and private messages, videos, images and group memberships. The social networking aspect lies in establishing connections between these digital manifestations of their identities. This creates a network of associations called friends who might have extra privileges with regard to the content being shared by them. Alternatively, these friends are simply listed on the profile pages to indicate a special relation. Linking profiles together is intended to display existing affiliations and peer groups. SNSs encourage users to exchange information about themselves, and to expand their network. [6]

According to a 2007 survey conducted by the Pew Internet & American Life Project more than half of all online American youths ages 12-17 use social networking sites [26]. Indeed, a great amount of SNS users are teenagers and young adults in their twenties, and as with all social interaction, they seek to explore themselves, relationships, social status and cultural norms; they post pictures of themselves and keep online journals. Much of what they share are cultural artifacts, such as, fashion and media. Exploiting the communication features of SNSs users keep in contact with old acquaintances, reinforce existing friendships, and by broadcasting their identity they may also meet and bond with new like-minded people [26, 23, 2]. On the other hand, networking skills are increasingly important in all aspects of today’s society and social networking services are also being used as professional tools to extend the network of work-related contacts. As SNSs have risen to public awareness and become mainstream, the age of the users has started to skew towards an older average [8].

The business value of SNSs is rooted in the unique collections of organized sets of preference information they generate. Users willingly spend great amounts of time to fine-tune the digital representation of themselves to best reflect their interests. In a 2007 survey conducted by the Pew Internet & American Life Project they found that “Almost half of social network-using teens visit the sites either once a day (26%) or several times a day (22%).” [26]. Furthermore, the preference information is likely to be accompanied by additional details, such as, age and marital status, which can be used to create intricate marketing models based on demography. Indeed, personal data has become a hot commodity and it is being repackaged and sold for many different purposes. Marketing and ad agencies, political organizations and employers, to mention but a few, all stand to gain from harnessing this unprecedented view into the minds of consumers, voters and job applicants effectively.

SNSs themselves are, as well, a powerful platform for serving ads, able to deliver targeted material to focus groups with chosen attributes based on their profiles. Facebook uses the term social ads for its highly targeted form of advertising [13]. Being among the most visited sites on the Internet, SNSs are, indeed, an excellent venue for forging and promoting brands. According to a report by comScore Media Metrix, Facebook was sixth in list of web properties serving most ads in November 2007 in the US [10]. Finally, another source of revenue is charging for service personalization and customization capabilities. All in all, considering its pure business value alone, online social networking is not something that is going to go away. The fact that Microsoft bought a 1.6% stake in Facebook at the price of $240 million illustrates the degree of potential business value believed to reside in online social networking [36].

Privacy is “the right of people to control what details about their lives stay inside their houses and what leaks to the outside” [32]. It is the focus of concern when it comes to online social networking. There are many things we are not deliberately seeking to hide but would not want to be freely distributed either. As individuals we need privacy because surveillance information can be taken out of context and abused in several ways. This is why, citizens and consumers should know what information is collected about them and by whom.

In addition to privacy-related issues, social web applications are a fertile hunting ground for criminals looking to scam people. Functionality for social interaction and implicit or explicit trust networks provide them with new tools of deception to utilize in their malicious endeavours. In a 2008 report the Internet Crime Complaint Center 9 in the US reported an all time high monetary loss from Internet crime in 2007, amounting to $240 million [18]. The data also indicates that Internet fraud is increasing.

3 Threats and Solutions

SNSs are a mode of communication unlike any other. There are many specific properties that set them apart from the offline world. Computer-mediated communication is generally persistent, searchable and replicable, and the audiences are invisible [5]. What information you choose to disclose generally sticks around because of caching, replication and archiving, and you have little or no control over by whom, or for what purposes this information is viewed. Also, so-called spidering programs automatically scour the web and aggregate, index and categorize information for easy searchability. Weak password-based single-layer non-SSL authentication methods and access control may enable this automated data collection to also reach information expected to be protected; data whose access is restricted to a specified limited group of friends. One such incident surfaced in December of 2007 when Facebook accused aporn company of trying to collect information from its service [21].

9IC3 is a joint operation between the Federal Bureau of Investigation and the National White Collar Crime Center to serve as a vehicle to receive cyber crime complaints from private citizens and industry in the US (http://www.ic3.gov).
In an unmediated environment you can look around and see who might overhear you. Based on this and the reactions of the people present, you constantly adjust what you’re saying to fit the social context. Whereas, once information is published online, it should be expected to be available for anyone anywhere anytime. Finally, online communication entails an apparent loss of deniability with conversations intended to be private afterwards being made available to the public. Although, nowadays this very much applies to face-to-face interactions, as well, due to the ubiquity of mobile phones and other small digital recording devices capable of audio and video recording.

3.1 Threats

The threats fall in two main categories. The first group comprises mainly privacy issues specifically related to the social web: misuse of, misrepresentation of or unauthorized access to sensitive personal information. Besides this, there are many security-related threats that are already prevalent in more traditional media, such as spam and malware. Users of SNSs share a lot of information about themselves. They might wish to be easily identifiable to be found by their friends, or just hope to attract like-minded people. They do not seem to understand that the intentions of the other users might be completely different from theirs. In a 2007 survey conducted by the Pew Internet & American Life Project 40% of examined profiles viewable online were knowingly left or set visible to anyone. Only 1% did not know what their visibility settings were. In a survey of Michigan State University students registered on Facebook it was found that very few, below 10%, believed that their profile might have been viewed by MSU administration or law enforcement. In a related study it was found that only 19% of the profiles were set as private, that is to say, only viewable by friends. Moreover, they found that on average users fill out 59% of the information fields available disclosing much information and discovered a positive correlation between the amount of information provided, and the number of friend links. Others have had similar results: in an analysis of the online behavior of Carnegie Mellon University students it was found that “90.8% of profiles contain an image, 87.8% of users reveal their births date, 39.9% list a phone number, and 50.8% list their current residence.”

On the one hand, the presumption of security due to the lack of physical interaction, and on the other hand, the false sense of intimacy created by seemingly private conversations causes people to get caught up with sharing anything ranging from their schedule to youthful indiscretions. OSN can be very addictive with users trying to accumulate their network and in so doing lose any restrictions imposed on the so-called non-friends. All in all, the users often end up revealing a lot of information, such as, their real names, ages, locations, sexual preferences and political views. They openly discuss drug use and publish accounts of underage drinking. They seek attention by being provocative and do things they would never do offline. What once was written in a personal diary or talked through in a phone conversation or face-to-face is now being stored online.

This kind of indiscriminate information revelation renders the users vulnerable to bullying, stalking, harassment, identity theft, sexual predators and other abusive behavior. One driver of offensive behavior is also the apparent anonymity provided by the web, which makes people lose their sense of social responsibility. Bullying can, for example, take the form of defamation with profile squatting, where a fake profile representing some person is filled with insulting information. Stalking and harassment is made straightforward with people sharing the names of their friends, hobbies and addresses, their whole schedule. Falling victim to identity theft can, for example, result in large loans taken in your name.

Something very important to bear in mind is that while you might not intentionally reveal your identity or location, these could be inferred from your writings, images or other data you give out. For example, cameras can add in metadata to your images, such as the date and time when the picture was taken, GPS location, camera serial number or even a complete original thumbnail of the picture. Profiles can also be linked to you when other users you know tag images where you appear with your name and profile. The non-privacy threats related to SNSs are more or less the same as those commonplace in the traditional web and email: spam, cross-site scripting (XSS), malware, phishing, corporate espionage and fraud. Malware is especially made possible by the facilities for creating and distributing 3rd party applications on many social networking service platforms. Phishing, on the other hand, is an attack where by masquerading as a trustworthy entity in an electronic communication the attacker tries to obtain sensitive information such as credit card numbers. A common example is setting up fake bank sites and sending out messages asking the users to log on to them. Spear-phishing is a highly targeted form of phishing, e.g. aimed at an individual expected to be a lucrative or an easy victim. SNSs can be used to collect information for this kind of attacks in an attempt to set a more believable trap or choose good targets. SNS’ messaging capabilities can also simply be put to use as a channel for spam or in more elaborate schemes employing social engineering tactics to carry out scams, frauds or corporate espionage.

What is often left unsaid in discussions about the privacy and security issues of web 2.0 applications and SNSs is the undeniable fact that for many service providers there is not much incentive to work on these areas of their service. Ease of use and the bells and whistles stomp security when it comes to attracting lots of users. Also, to sell information for marketing and similar purposes, it is really in the sites’ interest to collect as much and as detailed data as possible, not to warn the user of disclosing too much information. To this end, the privacy policies and EULAs are often unclear on the usage and amount of data recorded, not to mention, complex, hard to read and subject to change at any time. In Face

\footnote{Identity theft refers to a fraud where by pretending to be someone else the perpetrator acquires money or benefits from this in some other way.}

\footnote{Global Positioning System}

\footnote{End User License Agreement, the terms to which the user must agree to use the software.}
book’s privacy policy they retain the right to collect information about the users from all possible sources: “Facebook may also collect information about you from other sources, such as newspapers, blogs, instant messaging services, and other users of the Facebook service through the operation of the service (e.g., photo tags)” [14].

In general, services do not provide adequate functionality to control privacy or disable it by default, and the terms require users to provide accurate, current and complete information. By accepting the dubious conditions users are willingly giving over control of their personal data. Perception is everything here, people believe they own their data until they, e.g. attempt to remove it [4]. Actions lack transparency and do not provide sufficient information about the risks and only reactively under a lot of public pressure due to user outrage do the service providers fix privacy infringements made possible by the service [34, 28, 30]. The sites lack proper security measures related to authentication, access control or protection from spidering, i.e. bulk data collection. Even if the site offers some control over access to the data, like the terms of use, the user preferences could be reset or changed anytime. Finally, the terms by which users’ details are given out to third parties are not well-defined.

Issues of privacy are not limited to abusive behavior. There are many entities that might find the personal information interesting in other ways. Something that seems perfectly fine to share online now and to a targeted group of friends, might not seem so wise to publish a few years later or to a different audience. Sharing the wrong kind of information can have unforeseeable consequences later in your life with e.g. employers performing online character checks to see what kind of life you lead. There is much what can be considered harmless until it ends up in the hands of a prospective employer, college recruiter, your insurance company or a relative. Some colleges have even expelled students for violating codes of conduct when they have come across photographs of underage binge drinking [7]. Finally, the government and the police monitor and record data, as well, which can be seen as a form of surveillance and an invasion of privacy. For example, the government could do finely targeted military recruitment based on data collected online.

3.2 Solutions

The core of the problem is that, whatever the reasons, people do not appear to care so much about the loss of their privacy and, on the other hand, it is not in the service providers’ interest to overly zealously safeguard privacy. To combat these concerns, some schools have officially banned or discouraged the use of SNSs [27]. However, for today’s teens this might be very much akin to taking away their friends. Apart from a complete ban, there are three main approaches: social, technical and legal solutions.

The first method of tackling the issue is raising awareness of it by educating people. There are already many sites with compilations of guidelines how to live safely in the online world [14]. However, naively discouraging people to not give out personal information is too simplistic an advice when it comes to the countless ways we interact online. There is a high risk of unknowingly revealing sensitive information. Everyone should actively guard their sensitive data, such as, locations, phone number and financial information.

Encouraging parents to monitor their children’s online behavior is all well and good too, however, those who are the most at risk are the ones already lacking any proper parent involvement. The main goal should be to get everyone to understand the possible consequences of making some information publicly available and really think through who might come into contact with it. Sitting alone at the computer might feel like a private exchange, however, the expectation should be that anything you publish on the Internet, whatever the intended audience, is publicly accessible and could be obtained by anyone. A good rule of thumb is to picture SNSs as public billboards. We all know that it might not be a good idea to just walk down the street and start telling about yourself to complete strangers, however, with most people this insight does not seem to transfer so well to the online environment.

Apart from dealing out sensitive information, the same as with malware and scams spread by email applies here too: think twice before opening links to steer clear of malware and be wary of criminals scanning for potential scam victims. [1]

The engineering approach is to develop and improve the technical tools of controlling privacy and security. For one thing, users should have more fine-grained control over what information they share with whom. This attempt at increased privacy, however, is easily made void by the addictive nature of expanding your network of friends, which by no means is discouraged, on the contrary.

To protect the younger generation the establishment of proper age verification might be the answer. The user experience would then be adapted according to the age. The centralized government approach of binding digital identities to real persons could as well solve many of the problems presented, however, there is always the question of the reliability of this identification method and the danger of identity theft and even worse privacy online when it comes to such entities as the government. This has been implemented in South Korea where each individual is assigned a unique resident registration number, which is commonly used for authentication in web services. It is, however, relatively easy to steal these identifiers, which has lead to obvious problems and an attempt to deter this kind of activity with strict legislation regarding illegal possession of these numbers [33, 12]. Half-way implemented security measures might even be damaging if they give an illusion of privacy where there is none.

Finally, the last resort is brute-force monitoring and filtering of content, ranging from manual screening to automatic examination, which have their obvious shortcomings of being very resource-intensive and therefore highly undesirable approaches from the service providers’ point of view. [1]

The final approach is a regulatory one. The aim would be to improve legislation to better address privacy and security issues in today’s web services. Privacy policies and EULAs should be independently reviewed and monitored and there
should be more strict requirements on their visibility and detail. Moreover, there should be more restrictions on what information is allowed to be collected and the data handling practices should be more transparent. It should not be possible to track and store just any data if the user consents to that, because users generally agree to almost anything lacking real grasp of the consequences. Above all, service providers ought to be obligated by law to implement all information features as opt-in, instead of the common prechecked box with a vague and potentially misleading description. However, the fact that laws vary widely between countries and the law-making process might be fairly slow makes the legislative approach very challenging if not impossible.

3.3 Future outlook

With the rise of context- and location-aware systems, and ubiquitous computing the data mining possibilities will only get worse as far as protecting privacy is concerned. Examples vary from the wearable bracelet-type device iBand [20] to GPS- and WiFi-enabled devices, such as mobile phones, which are already offering services making use of location data. Mobile social networking is still a niche activity but micro-blogging with applications such as 15, dodgeball17, jaiku18 and pownce19, is growing in popularity [37, 35, 22, 17]. Advances in face recognition technology will also make photo-sharing sites like Flickr a repository of sensitive information [31, 3]. At first, the sheer amount of data may be a small barrier of safety but with the ongoing development of the semantic web and the inevitable tera machines with such amounts of processing power and storage space that they could record every minute detail of our daily lives, this comfort will soon be gone. If that were not enough, cybercrime generally has high rewards but low risk of getting caught due to the difficulty of tracing and pursuing criminals over state borders.

4 Summary

As the outside world is perceived increasingly dangerous to children, and people generally spend more and more time online, they flock to the social web to come together and express themselves. It could be argued that all the technological sophistication we are seeing online is nothing new, but that the new medium just makes it more visible. There’s no denying, however, that the reality TV generation feels comfortable sharing many aspects of their lives the previous generations would have kept private. Nonetheless, as is evidenced by the recurrent uproars about privacy intrusions, this is much due to them not fully understanding the risks involved. But it will take many more victims before the message is received. In the time when Google and Facebook are verbs, it might already be too late for the current generation, who, perhaps, have lost their privacy for good.

All of the solutions presented in this paper have their flaws. The social approach of simply educating people leaves too much in the hands of the users who despite education might not be able to comprehend the full ramifications of their actions, and would still make the same mistakes. In the technical approach the burden is placed on the service providers who generally act on the laws of demand and supply, providing features what the mass public is asking for, which at least, for now, do not include advanced security and privacy controls. This and other problems could be alleviated with laws pertaining to privacy and security in online services but this is a challenging, if not an impossible, effort due to the difficulty and slowness of harmonization of differing laws.

Instead, we need a combination of approaches. Implementing all information sharing features as opt-in would definitely be a giant leap in the right direction, but that alone is not enough. Education is still needed, as the first step, to make everyone realize the very real risks involved. In the end, the best safeguard is, as always, critical thinking and educated common sense in managing the online identity and in navigating the online world. Much the same rules apply as in the real world. Furthermore, on a society level, privacy does not mean just keeping excessive information away from criminals but also from companies, governments and other similar entities. Educating people about the use and potential misuse of personal information is the key.

All the criminal by-products that are nowadays commonplace on the web and in e-mail will in time become similarly wide-spread in the new communication media, the SNSs, as well, as has been evidenced by some recent incidents [15, 29]. Frauds, malware, phishing attacks and spam will be reinvented and reformed to fit and make full use of the the online social networking context.

Parents face serious challenges in teaching their children to manage in these online environments. How are young children to cope when even adults are unable to recognize frauds and privacy infringements? On the other hand, it is yet unclear how the new ways of information flow will affect the society as a whole, it could be we are merely going through a painful transition phase to a different kind of information society.

5 Future Work

There is much research on what youngsters seek to gain from interaction on social networking sites but as the services have become mainstream the demographics are changing and the users are older than before. According to a report by comScore Media Metrix published in October of 2006 more than half of the MySpace users were 35 or older [8]. We need to look into what these new age groups use SNSs for and how, and what, if any, differences there are in their way of managing their online identities when it comes to privacy and security.

15 A form of blogging that allows users to write brief text updates, and publish them e.g. by means of text messaging, instant messaging or email.
16 http://twitter.com/
17 http://www.dodgeball.com/
18 http://jaiku.com/
19 http://pownce.com/
20 http://www.flickr.com/
21 http://www.google.com/
References


