Self-Imposed Violations of Privacy in Virtual Communities

John Warren
Department of Information Systems and Technology Management
University of Texas at San Antonio
One UTSA Circle
San Antonio, TX 78249-0634
Phone: 210-458-5387
john.warren@utsa.edu

Copyright ©2006 by the UTSA College of Business. All rights reserved. This document can be downloaded without charge for educational purposes from the UTSA College of Business Working Paper Series (business.utsa.edu/wp) without explicit permission, provided that full credit, including © notice, is given to the source. The views expressed are those of the individual author(s) and do not necessarily reflect official positions of UTSA, the College of Business, or any individual department.
Self-Imposed Violations of Privacy in Virtual Communities

John Warren
Department of Information Systems and Technology Management
University of Texas at San Antonio
One UTSA Circle
San Antonio, TX 78249-0634
Phone: 210-458-5387
john.warren@utsa.edu

ABSTRACT
Virtual communities have grown in recent years due to the accessibility and ease of setting up web pages on communal Internet sites. Organizations, including companies that seek to hire new employees, often scan these web sites as part of their background checking process. One of the most prominent sites is MySpace.com. This research investigates issues of privacy and self-violation of privacy in virtual communities. The results indicate that individuals willingly provide personal information that may actually violate their privacy.

Keywords: Virtual communities; privacy; violations of privacy; MySpace; Information Systems; online social networks.

JEL Code: M150

ACKNOWLEDGEMENT:
This research was supported by a grant from the 2007 Summer Research Program of the College of Business at the University of Texas at San Antonio, USA
1. INTRODUCTION

The growth of the Internet has contributed to the proliferation of social activities online. This vast computer network facilitates the creation and existence of online social networks, commonly known as virtual communities (VCs). VCs offer advantages to members of these networks by allowing them to interact across different time zones, at anytime during the 24-hour day, and they eliminate the necessity for members to meet face-to-face.

A social network, in general, is a relationship between people, drawn together by common interests. It has been defined as a social structure made of nodes that are linked by one or more specific types of relations (Cook, 2001). Nodes generally refer to individual points. Ties represent the relationships between two nodes. The strength of a tie can be determined by how often, within a prescribed period of time, that the two nodes have interacted. When a computer network connects geographically dispersed people for exchanging information, it is in effect a social network (Wang and Chen, 2004; Wellman, 1996). The participants in these networks are generally widely distributed, and the number of potential participants is large.

Social networks are important because they provide individuals with social capital. Social capital has been referred to as benefits that may accrue to groups and individuals consequent to social interactions (Coleman, 1988; Ellison, et al., 2007; Granovetter, 1973; Parameswaran and Whinston, 2007a; Woolcock, 1998). It allows individuals to draw on resources from other members. Access to individuals outside one’s close circle of friends provides access to non-redundant information, which may result in benefits such as employee connections or business opportunities (Ellison et al., 2007 Granovetter, 1973). Social capital comes about through the changes in the relations among persons.

Virtual communities are computer-mediated communities, “where people form a social network by sharing information and knowledge, achieving socialization, or making transactions (Wang and Chen,
2004). Porter (2004) defines the VC as “an aggregation of individuals or business partners who interact around a shared interest, where the interaction is at least partially supported and/or mediated by technology and guided by some protocols or norms.” These communities are social networks that often spans large distances, allowing member that may be separated by space and time to establish, build and maintain relationships with each other. Relationships with friends, relatives and former neighbors can be maintained even if the individuals are separated geographically. New ties can be developed among people that share similar interests but may have never met physically. These VCs permit time and space to become less important and makes it easy for individuals to communicate with large groups of community members.

In earlier times, individuals would often “sow their wild oats” during their youth. This was often considered to be somewhat of a “rite-of-passage” to growing up and part of the maturing process. Later on, they would put away these behaviors and often be none the worse for it. However, the growth of the Internet and particularly, VCs, has contributed to a permanency of these youthful indiscretions. In today’s society, posting personal information has become increasingly common. Participating in online communities where individual users’ actions can be tracked without the users’ awareness nor permission threatens the very principles of freedom and openness that the Internet was founded on. Thus, these previously considered “innocent “youthful behaviors can now follow one indefinitely and can impact important circumstances later in life. Protecting the personal privacy of those participating in VCs should be a major concern for users, designers of online VCs and academic researchers.

Although there are laws designed to protect the privacy of individuals, many individuals put themselves at risk for privacy violations by willingly posting personal and sometimes potentially damaging information about themselves online. This information that can be gleaned from the Internet is information that “we put out there to be picked up (Nesson, 2001).”
In today’s VCs, the information that is contained on users’ sites often has potentially embarrassing details that may affect their ability to land a job in the future. In a recent survey by CareerBuilder (http://www.digg.com/tech_news/Be_careful_what_you_write_online, 2005), one in four managers now ‘Google’ potential employees and 51% of applications were rejected because of what was found (Montermini, 2005, Mossavar-Rahmani, 200. When asked to divulge the types of information discovered on the Web that caused them to dismiss potential employees, hiring managers pointed to the following:

- 31% - candidate lied about qualifications
- 25% - candidate had poor communication skills
- 24% - candidate was linked to criminal behavior
- 19% - candidate bad-mouthed their previous company or fellow employee
- 19% - candidate posted information about them drinking or using drugs
- 15% - candidate shared confidential information from previous employers
- 12% - candidate lied about an absence
- 11% - candidate posted provocative or inappropriate photographs
- 8% - candidate’s screen name was unprofessional

A separate study conducted by the executive job-search agency ExecuNet found that 75 percent of recruiters already use Web searching as part of the applicant screening process. More than a quarter of these recruiters have eliminated candidates base on what they found online (Greenfield and Haugh, 2006).

McCrea (2007) reports that these VC sites are proving to be a vehicle for employers seeking the right job candidate. Employers are increasingly doing “subtle” background checks on these sites. Drun Associates, Inc., searches information available on a site like MySpace, keeping an eye out for anything
that might reveal the person’s character, or perhaps hinder his or her ability to perform reliably. “Sometimes all we find is meaningless chitchat,” ways Drum, “but once in a while we’ll turn up something useful, like an unflattering picture or a piece of information that really shows what the person is made of [McCrea, 2007].”

The main objective of this paper is to provide an overview on VCs and privacy issues specific to VCs. In our analysis, our goal is to determine the extent to which individuals place content on VC sites that may be damaging to them, thus violating their privacy. We also discuss how organizations can provide intervention that may increase awareness and minimize some of the potential damaging behaviors exhibited by individuals on these sites.

This research adds to the literature on privacy as it pertains to virtual communities. Self-imposed violations of privacy in VCs have not been adequately addressed in previous literature. With the arrival and proliferation of the Internet, research focus in information systems should address these emerging trends, including VCs (Parameswaran and Whinston, 2007a; Parameswaran and Whinston, 2007b).

This paper is organized as follows. First, we provide a discussion of the existing literature on privacy and privacy in virtual communities. Second, we discuss one of the most popular virtual communities that exist on the Internet. Third, we provide and discuss the methodology for a study involving the polling and analysis of 1104 MySpace web sites. Fourth, an analysis and the results of the data collection are presented. Last, a summary discussion and conclusions are presented.

2. PRIVACY AND VIRTUAL COMMUNITIES

2.1 Privacy

Privacy, in general, can be defined as the need to secure for the individual “the right to be left alone” or as the “state or condition of limited access to a person” (Schoemann, 1984, Turner and Dasgupta,
A more classic definition of privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others (Turner and Dasgupta, 2003, Westin, 1967). Awad and Krishman (2006) define privacy as “the ability of the individual to control the terms under which personal information is acquired and used” (see also Westin, 1967). In terms of VCs, an individual’s privacy or lack thereof may be determined by the information that the individual willingly posts on their sites. The threat to privacy has to do with the way that information posted is or can be used by others. An important question that individuals participating in VCs should consider is: are you in control of the privacy of the data that “you” post to your VC web pages?

2.2 Privacy on the Internet

There are several ways in which the Internet poses privacy threats to users. The collection of personal information via the Internet is fast and easy to do (e.g., web crawlers). Thus, personal information, once posted on the Internet, can be easily collected, stored, and retrieved, without the knowledge of individuals. Once this information has been collected, the dissemination of the information is rather easy. The decontextualization of personal information can also be a problem. Information that may be collected, archived and possibly traded may be used in another totally different context from which the originator of the information intended (Goldie, 2003).

Privacy is often discussed in terms of security. However, these concepts can be different, depending on the context in which they are discussed. Generally speaking, security has to do with making sure that any personal data provided on the Internet is transmitted and stored in a safe and protected way. As mentioned above, privacy has to do with the way that information posted or stored on the Internet is being used or has the potential for being used by others. These concepts are interrelated in that violations of privacy are often considered to be problems associated with security deficiencies.
2.3 Privacy in Virtual Communities

Goldie (2003) suggests that part of the appeal of the Internet has to do with its ability to connect people, unbounded by time and space restraints. Individuals meet online in various forums to discuss various issues that are often personal. Because these discussions are often personal and central to one’s identity, privacy risks are present. Individuals may discuss personal matters online that can easily be accessed by others, and may be archived indefinitely. Future employers, police investigators, and others, could potentially use the information in the online postings and discussions, thereby representing a threat to the privacy of VC members. All Internet users face privacy risks, but the risks for VC participants are unique.

Many VCs are set up so that the users’ identities are not readily apparent. However, often there is enough information provided to identify the site’s owner. Face re-identification has been discussed as one way that users may be identified. Users often re-use the same or similar photos across different sites, resulting in an “identified face” that can be used to identify a pseudonym profile with the same or similar face on another site (Gross, 2005, Gross and Acquisti, 2005, Liu and Maes, 2005). Other information that users often provide includes placing their actual names in the web address for their sites. They often post personal information such as the cities they reside in, the schools attended, place of occupation, and other personal information. This information is often provided during the initial web page setup by filling out a convenient and easy-to-complete template. Even if users do not provide their actual names, their names can often be discovered by reading their web logs, where their friends may address them by name.

Why do people place content on the Internet that may prove to be damaging in the future? Extant research suggests that the concern for information privacy is very high among consumers (Stark and Hodge, 2004). However, social networking sites, which facilitate the exchange of personal information, are booming in popularity. This seems to be a paradox. If consumers are concerned about privacy, what makes them willing to disclose information in social networking sites? One reason may be that Internet
users usually trust strangers, much like people gave rides to hitchhikers in the flower child days of the 1960s (Wellman and Gulia, 1997). Another reason may be that users often express very strong concerns about privacy of their personal information, but are less vigilant about safeguarding it (Awad and Krishman, 2006, Dwyer et al., 2007).

We believe that social capital theory may explain why people sometimes place potentially damaging personal information on the Internet. Social capital has been defined as “norms and networks facilitating collective actions for mutual benefits (Parameswaran and Whinston, 2007a, Woolcock, 1998). It implies that individuals engage in interactions that may lead to their benefit, despite knowing that there are risks of undesirable outcomes as a result of their interactions. Dwyer [2007] found that although individuals are concerned about the privacy in online social networks, they typically accept the tradeoff of access to no-fee sites in exchange for diminished protection of their private information.

Individuals may have a certain tolerance for risk in order to reap the benefits of posting personal information online (e.g., making new friends or trying to promote a business), thus increasing their social capital. The perceived benefits of revealing data to strangers may be larger than the perceived costs of potential privacy invasions. Other reasons include relaxed attitudes toward personal privacy, incomplete information about the consequences of willingly positing personal information, faith in the networking service, or a myopic evaluation of privacy risks (Gross and Acquisti, 2005).

VCs often have the feature in their users’ profiles where the users can set their profiles to private. This should allow the users to control who has access to their sites. However social engineering -- techniques used to manipulate people into performing actions or divulging confidential information, is a well-known practice in computer security in which individuals obtain confidential information by manipulating legitimate users (Mitnick et al., 2002). In popular online social networking sites such as MySpace and Facebook, all that a user has to do is to ask to be someone’s friend. Even though most users do not know the intentions of the person asking to be their friend, a surprising high rate of individuals will
accept the request (Gross, 2005, Jump, 2005). Once the user becomes a friend, the new friend has total access to the site. Nesson [2001] indicates that “when it comes to privacy from electronic intrusion, the costs of protection are often higher than the perceived benefits (2001, p. 109).”

3. MYSPACE

MySpace ([www.myspace.com](http://www.myspace.com)) is a popular web site which facilitates social networking among its users. Founded in August 2003 (http://en.wikipedia.org/wiki/MySpace, 2008), it has grown to more than 100 million member profiles. Setting up a MySpace account is very easy, and consists of simply filling out an online form. After an individual signs up, he or she will have access to many tools that facilitates setting up the web site. Web building tools allow individuals to create web pages without needing to know how to program in HTML or how to use commercially available Web building software; you just fill in the blanks at MySpace to create your page. There is also an online photo album where you can post your profile picture and upload other photos that are viewable by other MySpace members. There is a built-in tool for providing web diaries, or blogs. MySpace allows individuals to make their blogs publicly viewable or viewable only to a select audience. Users can also use the video hosting tool that allows them to post and share short video clips.

Teenagers and young people in particular, view MySpace as a fun and entertaining way to meet new friends and to keep in contact with old friends that they may have separated from due to relocation based on various reasons (e.g., family relocations, young people going away to attend college). Participants often display music and videos and display their talents. Without VCs such as MySpace, individuals may run the risk of losing social capital after a job related or college relocation.

Some of the positive features of MySpace are the following:

- It’s free
- Individuals can meet people all over the world and keep in touch
• It’s easy to use – the program and templates guide you in the web page creation
• Individuals can generate codes for the page
• It is a good network for business and music
• You can easily place your music files on the MySpace site for others to listen to
• People can leave messages (in the blogs section) for all to see
• Individuals can leave private messages
• Individuals can post short video clips
• Online surveys can be conducted
• Member profiles are posted on the page for people to get familiar with
• Individuals can search for members based on names, email, user name, or use the advanced search features
• If the age of the applicant is younger than 14, the system will delete your name automatically for safety reasons
• There are safety tips and tips for parents
• The police have been able to intercept and stop fights, violence in schools, graffiti, etc.

Negatives features include:

• It is easy for people (e.g., sexual predators, stalkers) to find others (physically) if they put too much personal information on the site.
• Most schools (k-12) block MySpace sites from students in school computer laboratories.
• It is easy to get addicted to which can result in the waste of time.
4. METHODOLOGY

To investigate the self imposed violations of privacy in VCs, we conducted a study of MySpace web sites. A total of 1104 MySpace web sites were randomly polled during the summer and fall of 2006. The purpose of polling is to get information from a few people and use it to learn about the larger population. A randomly selected, small percentage of a population of people can represent the behavior of the larger population (Utts, 1999). Sample size is dictated by how accurate you must be, or how large a margin of error you can tolerate. Utts (1999) indicates that you can usually estimate the margin of error by finding the square root of the sample size, then dividing 1 by that number (e.g., $\frac{1}{\sqrt{n}}$). A sample size with a margin of error of 3% or less is considered to be good. Our sample of 1104 subjects yielded a margin of error of 2.97%.

Data collection for the 1104 sites took over 6 months and approximately 400 hours. The average time per site was approximately 20 minutes. The advanced browse features of MySpace were used to complete the searches. This allowed the searches to be conducted randomly in different states throughout the United States. An example of the advanced browse screen is shown in Figure 1.
The types of data collected from the sites are provided in Table 1. Additional time was consumed cleaning data, elimination of duplicate site data and getting the data ready for analysis. We did not consider any of the profiles that were set to private because it would take too long to wait until the site owners added the researcher as a friend. The sites polled included 51 states and 17 countries.
Table 1. Data Collected from MySpace User Web Sites

<table>
<thead>
<tr>
<th>Category</th>
<th>Options</th>
</tr>
</thead>
<tbody>
<tr>
<td>URL</td>
<td>Enter the url for the site</td>
</tr>
<tr>
<td>First name posted</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Last name posted</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Photo posted</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Age/Birthdate</td>
<td>Enter the age</td>
</tr>
<tr>
<td>Gender</td>
<td>Male or Female</td>
</tr>
<tr>
<td>Phone number posted</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Street Address posted</td>
<td>Yes or No</td>
</tr>
<tr>
<td>City</td>
<td>Enter the name of the city</td>
</tr>
<tr>
<td>State or Country</td>
<td>Enter the name of the state or country</td>
</tr>
<tr>
<td>Other Address</td>
<td>Yes or No</td>
</tr>
<tr>
<td>School Name</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Company Name</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Marital Status</td>
<td>Married; single; divorced</td>
</tr>
<tr>
<td>Sexual Orientation</td>
<td>Straight; bi-sexual; gay/lesbian</td>
</tr>
<tr>
<td>Education</td>
<td>In grade school; completed grade school; in high school; high school graduate; in college; college graduate; graduate school or higher</td>
</tr>
<tr>
<td>Occupation</td>
<td>Enter occupation if given</td>
</tr>
<tr>
<td>Ethnicity</td>
<td>Caucasian; African American; Hispanic; Native American; Pacific Islander; Asian; Indian; Other</td>
</tr>
<tr>
<td>Religion</td>
<td>Enter the religion</td>
</tr>
<tr>
<td>Drink alcohol regularly</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Smoke</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Ever use Drugs</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Trouble with the law</td>
<td>Yes or No</td>
</tr>
<tr>
<td>Sexual promiscuity</td>
<td>Yes – Photos; Yes- Comments; No</td>
</tr>
<tr>
<td>Business advertisement</td>
<td>Yes or No</td>
</tr>
</tbody>
</table>
5. ANALYSIS AND RESULTS

The sample population was randomly collected from 1104 MySpace sites. Among the individual web sites, there were 48% males, and 51% females, with an age range of 14-78 years and a mean age of 25 years. The demographic profile including ethnic breakdown is provided in Table 2.

Table 2. Demographic Profile

<table>
<thead>
<tr>
<th></th>
<th>Number (%)</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Sex</strong></td>
<td></td>
</tr>
<tr>
<td>Male</td>
<td>534 (48.360%)</td>
</tr>
<tr>
<td>Female</td>
<td>570 (51.63%)</td>
</tr>
<tr>
<td><strong>Age</strong></td>
<td></td>
</tr>
<tr>
<td>Range</td>
<td>14 – 78 years</td>
</tr>
<tr>
<td>Mean age</td>
<td>24.98</td>
</tr>
<tr>
<td><strong>Ethnicity</strong></td>
<td></td>
</tr>
<tr>
<td>Caucasian</td>
<td>521 (47.19%)</td>
</tr>
<tr>
<td>African American</td>
<td>188 (17.03%)</td>
</tr>
<tr>
<td>Hispanic</td>
<td>118 (11.69%)</td>
</tr>
<tr>
<td>Asian</td>
<td>55 (4.98%)</td>
</tr>
<tr>
<td>Native American</td>
<td>8 (.72%)</td>
</tr>
<tr>
<td>Pacific Islander</td>
<td>14 (1.23%)</td>
</tr>
<tr>
<td>Indian</td>
<td>2 (.18%)</td>
</tr>
<tr>
<td>Other</td>
<td>25 (2.26%)</td>
</tr>
<tr>
<td>None Given</td>
<td>173 (15.67%)</td>
</tr>
</tbody>
</table>
Of the 1104 sites that were studied, the majority of the site owners had photos of themselves posted (97%), listed at least their first names (86%), and listed their ages (98%). We did note that a few of the web site owners gave false ages. An additional 15% listed their last names along with their first names. Less than 1% listed their personal telephone numbers. 65% listed the school attended or currently attending, 26% listed the name of the business that they work for or own, 30% indicated alcohol use and drinking as activities that they enjoy on a regular basis, and 32% indicated that they were smokers. Only 3% indicated drug use and 14% were considered to be sexually promiscuous as evidenced by either photos or comments. 1% of the site owners actually indicated that they had criminal records.

Figure 2. Social Behavior of MySpace Users
5. DISCUSSION

The findings from our data collection indicate that users post information on sites that may be potentially damaging to them in the future. Information that “friends” post on blogs may reveal information that violates privacy (e.g., full name, school or work location, etc.). Even if users’ names were not listed on their websites, they can often be found by searching by name or email address. Many users believe that information posted on their MySpace pages are private and are viewed only by their friends (Gross and Acquisti, 2005, Jump, 2005).

One of the features of VCs that may facilitate users’ providing personal information has to do with the setting up process. In many social networking sites, during the registration process the users are asked personal questions about themselves, and they generally will answer them. The questions include things like their name, age, where they live, and questions regarding their friends. The threats comes from the fact that others can harvest this information using software tools, and potentially use this information to the detriment of the site owner.

It is difficult to manage and protect information that people are encouraged to disclose on VC web sites in order to fully benefit from what the sites offers. As a result, users should be aware of the vast quantity of information that is captured in the information technology systems that facilitates these sites. Most people are not fully aware as to the level that their behavior can be tracked when they are online. Users may have a false sense of security when participating in VCs. Although blogging on these sites is often easy and convenient, their activities can possibly lead to actions that may have legal consequences down the road (e.g., disclosing confidential information, breaking laws, revealing more information than users intend to). Third, there is a persistency of the information that is digitally captured. For example, postings of opinions run the risk of being discovered by audiences that it was not intended for (e.g., future employers) long after the initial postings (see Nabeth, 2007).
If we had gained access to “private” sites the results may have yielded more violations of privacy. Efforts to do this would have been time consuming. However, Dwyer (2007) noted that while most VC sites did offer the option of setting the site to “private,” most of the individuals did not choose to use this option. This is understandable because most people who set up web sites in VCs do so with the objective of letting people know who they are and what they do. They may want to promote their skills, develop and grow a business clientele, or just find new friends. Setting their profiles to private would defeat these goals.

Some of the limitations to this study include the realization that MySpace is a very large VC with well over 100 million registered users. It can actually be considered to be a VC that contains subsets of VCs. Although our margin of error was less than 3%, we still had a relatively small sample size. In addition, the postings on VC sites are not always accurate, but they are still potentially damaging. If potential employers, universities, predators, ext. read it, they are more likely to accept the postings as true than false.

Future research can include the addition of survey questionnaires to members of VCs along with the analysis of the individual sites. Also, the “unique” daily interactions in the VCs may provide additional insight. Last, the level of individual participation in “specialized” VC sites (e.g., online medical support groups) can be studied to try to determine the effects of participation in such sites. Many of these sites have registered users who do not participate in discussions and may be considered “free riders”, but they may contribute to the life span of the VC simply by the fact that they are registered users.
REFERENCES


