ABSTRACT

Social networking sites play an ever increasing role in people’s lives today. For three of the currently most popular and heaviest used social networks, i.e. Facebook, Google+ and Myspace, the sign-up process and posting processes were systematically analysed. A vulnerability metric was defined in order to be able to compare the results of this analysis, with the possibility to compare more additional social networks in the future. From these three social networking sites, Facebook is the least vulnerable regarding the sign-up process and Google+ is the least vulnerable regarding new postings. Myspace is the most vulnerable regarding both social networking processes.

Keywords

Social networks, private data, data protection, accessibility, vulnerability

1. INTRODUCTION

The last decade has witnessed the rise of social networks, which offer users the possibility to interact with each other over a variety of subjects. The definition of a social network is provided in many documents. In [1], for example, social networks are defined as “web-based services that allow individuals to (1) construct a public or semi-public profile within a bounded system, (2) articulate a list of other users with whom they share a connection, and (3) view and traverse their list of connections and those made by others within the system.”

In general, social networks are platforms where people can create an online presence to connect with friends and other acquaintances (see Figure 1). Furthermore, many social networking sites offer the possibility for their members to subscribe to groups. These groups (which can have different names in different networks) can be about a wide variety of subjects. From the equivalent of an offline social group (e.g. a sports team or fraternity) to world-wide interest groups in a very particular subject (e.g. scientists in a certain specialized field of study) to groups with a broad range of followers (e.g., people who watch the series The Big Bang Theory).

Especially this last type of group offers an interesting opportunity to businesses and brands. Therefore, some social networking sites offer special possibilities to these venues to create an online presence on that social network.

Typically, the content on social networking sites is created by people or other presences such as companies posting updates to the network. These updates will then be aggregated and presented to everyone with an interest (e.g., friendship, follower, subscriber) to the updater.

Updates can consist of a wide variety of multimedia. Examples are blogposts, photos or other images, reactions to other people’s postings and links to articles outside the social network. In addition, social networks offer one-click buttons which users can click to show their affinity with a posting.

A relatively new aspect of social networking sites is their presence on mobile devices. Social network usage is not restricted to webbrowsers. For example, research shows that smartphones attest for more than 50 percent of Facebook use [3]. With data constantly being synchronised between mobile devices and social network backends, it is easier than ever to be in touch with others via social networks.

Along with the positive benefits that social networks can offer, for example in maintaining contact with private (facebook.com), professional (linkedin.com) or academic (academia.edu) acquaintances, researchers and human rights groups have been warning the public about the negative sides of all the self-disclosing possibilities that social networks entail as well.

Since it is very easy to post updates to social networking sites, people provide them with vast amounts of data. Primarily, this data is meant to be visible to the contacts a user has. Secondly, all data collected and connected to a specific user is a valuable source for data mining. Social
networks use this data to create user profiles, which they use for personalised advertising. This in turn, is the main source of income for social networks and the reason of why they offer free membership.

This paper however, focuses on a third effect that is induced by users posting vast amounts of information. Social networks employ a certain degree of openness of user profiles to people who do not have a connection with that user (yet). These other people do not necessarily need a profile to see some information of existing users. This can for example be useful to non-subscribers, since they are able to identify friends who are subscribed to the social network.

On the other hand, because of the degree of how personal some information can be, it can be possible that data that users rather keep and consider as private, becomes public. To study whether this is merely a hypothetical problem, or a real world situation, this research is performed.

This paper focuses on providing an insight on the accessibility and vulnerability of different social networks. In particular, this research activity investigates and attempts to devise a metric for the analysis of the accessibility and vulnerability of the data being disclosed by users of social networking sites. This metric could then be used to make (1) a comparison between different popular social networks, (2) a ranking of these social networking sites based on how they handle private data.

The main research question that is investigated by this research is: “How accessible and vulnerable is sensitive private data in popular social networks?”. Based on this main research question, several research questions are derived:

1. Which are the most popular social networks?
2. What accessibility settings are they offering to users?
3. Which private data makes users vulnerable under certain by default configured accessibility settings?
4. How could social networking data be classified/ranked from the point of privacy and vulnerability?
5. Which vulnerable private data fields of subscribers are by default accessible to other subscribers and to non-subscribers for different social networks?
6. Is there a difference between various social networks from the point of privacy based ranked social network data and the by default configured accessibility settings?

Research questions (1), (3) and (4) are answered by a literature study. Furthermore, research questions (2) and (5) are answered by accomplishing a data analysis based on the sampled social network private data fields. Moreover, research question (6) is answered by accomplishing a qualitative comparison and analysis.

This paper is organised as follows:

Section 2 provides an overview of the most popular social networks and their accessibility settings and answers research questions (1) and (2).

Section 3 provides a definition of private data and a linking of private data to accessibility settings. This section answers research questions (3), (4) and (5).

Section 4 provides a comparison of the introduced social networks based on the privacy and vulnerability of social network data. This section answers research question (6).

Section 5 lists the conclusions and gives recommendations for future work.

2. EXPLORING SOCIAL NETWORKS

Internet provides worldwide connectivity and online communication between people. Before the world wide web, the main means of meeting other people online was in the form of news groups. Web-based forums were used after the introduction of world wide web. In the last decade these web-based forums have evolved into social networks. Wikipedia [16] provides a non-exhaustive list of 206 notable and well-known social network sites.

This research will focus on three highly known and heavily used social networks. The selection of these three social networks is based upon [14]. These social networking sites are:

- Facebook
- Google+
- MySpace

Each social network provides various visibility options for the data provided by the social network subscribers. These visibility options are denoted as accessibility settings.

Due to the fact that different social networking sites use different terminology for their accessibility settings, the following five general terms for accessibility settings are introduced in this paper:

- **Public**: This accessibility setting represents the situation that the data provided by the social network subscribers can be viewed by everybody that can use an internet connection.
- **Members**: This accessibility setting represents the situation that the data provided by the social network subscribers can only be viewed by others who registered, i.e., are subscribed, with the same social network.
- **Contacts**: With this accessibility setting, the data provided by the social network subscribers can only be viewed by other registered subscribers of the same social network and who are authorised to do this by the owner of the data.
- **User**: Using this accessibility setting, the data provided by the social network subscribers can only be viewed by the subscriber that owns the data.
- **Custom**: This accessibility setting applies when a social network subscriber has the option to select specific custom subgroups that could access and view the data owned by the same subscriber.

The generic accessibility settings (Public, Members, Contacts and User) are ranked from the most to the least accessible. This means that someone that is authorised to access e.g., the Members setting is also authorised to access the Public setting.

The following subsections introduce the three social networks and their accessibility settings. Furthermore, the
accessibility settings offered by each social network will be mapped to the generic accessibility settings discussed previously in this section.

Additionally, for each social networking site it will be described which of the supported accessibility setting is the default accessibility setting for a posting made by the subscriber on that social network.

It is important to note that the discussed social networks do not necessarily offer the whole list of generic accessibility settings discussed previously in this section.

2.1 Facebook

The first social network that will be investigated is Facebook [7], the current epitome of social networking. Facebook was founded in early 2004 by then Harvard student Mark Zuckerberg.

Facebook can be considered a general purpose social networking site, where anyone can register. Apart from regular user accounts, other special type of accounts can be created by e.g., companies, celebrities and other types of brands. For example, United States of America President Obama and the University of Twente use such special accounts.

An important feature of Facebook is the so-called timeline. This feature supports the activity where all updates with regard to a user are being posted chronologically, with the ability to scroll back to earlier dates. Another feature is apps, which represent applications within Facebook used for the realisation of specific tasks/activities, e.g., Photoalbums and (social) games. These apps can be highly integrated in the Facebook platform.

Facebook is widely known for its ‘like’-button. This button can be added to media, such as news articles or blog postings, both on and off Facebook itself and can be used by a user as “an easy way to let someone know that you enjoy it, without leaving a comment.” [5]

Currently, Facebook ranks as the number one [14] social networking site, with over 900 million unique visitors in March 2012 [6].

In principle, both subscribers and non-subscribers can access Facebook pages. However, Facebook lets users decide to what extent they want to open their profile to others, by providing different accessibility settings, which are configurable per data field. Possible settings are displayed below. Each accessibility setting is mapped into one of the generic accessibility settings defined previously in this Section. These equivalent generic accessibility settings are below written in italic.

- **Public**: As the name suggests, this accessibility setting maps to the *Public* generic accessibility setting.
- **Friends**: Within Facebook, subscribers can be ‘friends’ with other subscribers. This setting corresponds with the *Contacts* generic accessibility setting.
- **Only Me**: This is most private accessibility setting supported by Facebook, which maps to the *User* generic accessibility setting.
- **Groups**: Facebook offers the possibility to create custom groups of people, such as ‘Friends of friends’. This setting maps to the *Custom* generic accessibility setting.

If a user makes a new posting to Facebook, he or she has the option to select any of the four accessibility settings mentioned above. However, the default accessibility setting that is used during posting is *Public*.

2.2 Google+

The second social networking site which will be investigated is Google+ [8] (pronounced Google Plus). After previous attempts with services named Orkut and Google Buzz, Google tries to get a foothold in the world of social networking by creating the Google+ social network [13].

Google+’s date of launch was the 28th of June, 2011. Initially it was only accessible to people who already had a Google account and had received an invitation from others who already had a Google+ account. Ninety days later, September 20, 2011, Google+ opened its doors to anyone (above a certain age which is country dependent) [9].

According to [14], Google+ is the fifth largest social networking site. On their official blog, Google claimed to have a user base of 170 million users in April 2012 [10].

To create a Google+ account, one needs to have an account with Google, which by default is a @gmail.com-account. For this research it was assumed that a user who is creating an account doesn’t have a Google account yet. Just like Facebook, Google+ offers different accessibility settings to data fields. Again they are mapped to the *italic* generic accessibility settings defined previously in this Section:

- **Public**: This setting maps to the *Public* generic accessibility setting. This accessibility setting is used in Google+, but as well in other Google products, including search.
- **Your circles**: In Google+, users organise their contacts in so-called circles. This is comparable to mathematical set theory, where a circle represents a set and the users contacts are elements of this set. This setting maps to the *Contacts* generic accessibility setting.
- **Only you**: This setting refers to private data, and it maps to the *User* generic accessibility setting.
- **Custom**: By using this setting, data can be made accessible to a combination of different circles and individual contacts. Therefore, this setting can be mapped to the *Custom* generic accessibility setting.

For making a new posting, Google+ offers *Public, Contacts and Custom* as possible accessibility settings for that posting. By default, the accessibility setting that is used during posting is *Contacts*.

2.3 MySpace

The third social networking site is Myspace [11], a social network mainly aimed for people to follow artists they like and share them with their friends.

Myspace was originally founded in 2003 and has for a while, between 2005 and 2008, been the most visited social network. In 2008 however, it was surpassed by Facebook and ever since there has been a decline in the market share of Myspace.

Still, [14] ranks Myspace as the fourth biggest social networking site, with slightly more monthly visitors than Google+.

When signing up for a Myspace account, a person has two options. The first option, which was used in this research, is the regular registration option where a person has to fill
in a certain amount of fields. The other option is to use Facebook credentials to sign up and log in.

Artists have the possibility to create specific artists accounts. When signing up, it is possible to choose between a personal, a musician, a comedian or a filmmaker account.

After registration, a user has the possibility to alter the accessibility settings for his or her account. The possible accessibility settings and how they correspond to the generic accessibility settings defined earlier in this Section are described below:

- **Everyone**: This accessibility setting maps to the Public generic setting.
- **Only my friends and anyone over 18**: This setting maps to the Members generic accessibility setting, because any member can easily change his or her age to 18 or higher.
- **My friends only**: Myspace uses a friends-system, where users can befriend other users. This accessibility setting corresponds to the Contacts generic accessibility setting.
- **No one**: This data can not be viewed by other. Therefore, this setting can be mapped to the User generic accessibility setting.

When a Myspace subscriber makes a new posting to the social network, the possible accessibility settings are Public, Members and Contacts. This is, however, not customisable per posting. Moreover, the default accessibility setting that is used during posting is Public.

3. **RANKING SOCIAL NETWORKING DATA FIELDS**

This section ranks the possible data fields maintained by the three selected social networks, based on their privacy and vulnerability. This ranking process is accomplished using the following approach.

The first step encompasses the sign-up process. Starting with no presence at each network, an account will be created. Per sign-up process, it will be administered which data fields need to be filled in and whether this filling in activity per data field is mandatory or optional.

Subsequently, after the account is created, it will be assessed which data fields are available to others. This will both be assessed from the perspective of other social network subscribers and as well from the perspective of non-subscribers. This will define whether the data fields can be considered private or not. In addition to this, each data field will also be graded as being vulnerable or not. Furthermore, for each social network it will be observed to what extend the accessibility setting can be adjusted.

Several definitions for privacy and vulnerability can be found in the literature. Regarding privacy, for example, Article 12 of the Universal Declaration of Human Rights, see [15], protects and defines privacy as the right of humans to keep information about themselves secluded from the public and to stay in control about information related to their personal affairs.

Furthermore, in European law, with relation to privacy, personal data is defined as "any information relating to an identified or identifiable natural person (‘data subject’); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity" [4].

In this paper, private data is defined as the data associated with a person and (1) is related to its personal affairs, and (2) can be used to identify the person with a little effort. Examples of private data are: family name, address, date of birth, email address, age, weight, height.

Regarding vulnerability, several definitions can be found, e.g., [2] gives the following definition: “Degree to which people, property, resources, systems, and cultural, economic, environmental, and social activity is susceptible to harm, degradation, or destruction on being exposed to a hostile agent or factor”.

Vulnerability can be applied on different social networking activities. In this paper it is applied to the notion of vulnerability in the sign-up and posting activities.

Vulnerability in the context of the sign-up social networking activity is associated with the used data fields and is defined as follows.

A data field is vulnerable when this data (1) is considered to be private data, (2) is easily accessible by non-subscribers and (3) when the social network site mandates its users to fill in this data.

Vulnerability in the context of the posting social networking activities is associated to the situation that a social network subscriber is posting data within the social network and is defined as follows.

A posting activity is rated as vulnerable depending on the values of the default accessibility settings introduced in Section 2.

A social network is considered to be the most vulnerable in the context of posting when the default post accessibility setting is Public. The social network is least vulnerable in the context of posting when the default post accessibility setting is User. Furthermore, the post accessibility setting Contacts is more vulnerable than User. Moreover, the post accessibility setting Members is more vulnerable than Contacts.

For each of the investigated social networking sites, a table is presented that provides an overview and classification of the data fields collected during the sign-up process. In particular, this table describes whether each of the supported data fields is (1) mandatory to be filled in, (2) using a default accessibility setting (D.A.S.), (3) private and (4) vulnerable. The elements used in each table are described below:

- **Data field**: This element provides the name and description of the data field requested by the social networking site.
- **Mandatory**: This binary (Yes or No) element represents whether the social network considers the data field as being mandatory to fill in during registration.
- **Default Accessibility Setting (D.A.S.)**: This element states how accessible the data field will show up after registration, assuming that the user did not alter it. There are five possible values (Public, Members, Contacts, User and Custom), which are described previously in the introductory part of this section.
- **Private**: Similar to the Mandatory element, this is also a binary element, where values can be equal
Table 1. Overview and classification of data fields during Facebook sign-up process

<table>
<thead>
<tr>
<th>Data field</th>
<th>Mandatory</th>
<th>D.A.S.</th>
<th>Private</th>
<th>Vulnerable</th>
</tr>
</thead>
<tbody>
<tr>
<td>First name</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Last name</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Email address</td>
<td>Yes</td>
<td>Contacts</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Date of Birth</td>
<td>Yes</td>
<td>Custom</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Sex</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Windows live id</td>
<td>No</td>
<td>User</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>College/University</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>High School</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Employer</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Profile picture</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Friends</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
</tbody>
</table>

• Mandatory: This item must be provided during the registration process.
• D.A.S.: Default Accessibility Setting
• Private: This item is private, i.e. only the person that registers knows about it.
• Vulnerable: This element, which again is a binary element, describes whether the data field is considered vulnerable, as described previously in this section. This occurs when the elements Mandatory, Default Accessibility Setting and Private, respectively have the values of Yes, Public and Yes.

3.1 Facebook
The Facebook sign-up process consists of multiple steps during which the user has to fill in a few forms. An overview and classification of the data fields collected during the Facebook sign-up process can be seen in Table 1.

As can be seen from Table 1, there are seven data fields which are considered as being private. Furthermore, a total of three data field which are encountered during the Facebook registration process are considered as being vulnerable: *First name*, *Last name* and *Sex*.

When a user makes a posting to Facebook, this will by default be done with the accessibility setting *Public*.

3.2 Google+
When signing up for a Google+-account, the user has to select whether or not he or she already has an account with Google, e.g. a @gmail.com-account. In case the user has such an account, less sign up information is required, since Google is already aware of this information.

In this research, it is considered that the user does not have an account with Google. This was done in order to use the same assumptions (i.e. no prior knowledge of the user) for all social networking sites.

During the sign-up process, a variety of data fields is requested by Google. An overview and classification of the data fields collected during the Google+ sign-up process is given in Table 2.

As can be seen from Table 2, there are seven data fields which are considered as being private. Furthermore, a total of three data field which are encountered during the Facebook registration process are considered as being vulnerable: *First name*, *Last name* and *Sex*.

When a user makes a posting to Facebook, this will by default be done with the accessibility setting *Public*. **Possible Friends:** This element suggests an algorithmically selected list of Facebook users, who have a high chance to be a real life friend of the person that registers.
### Table 2. Overview and classification of data fields during Google+ sign-up process

<table>
<thead>
<tr>
<th>Data field</th>
<th>Mandatory</th>
<th>D.A.S.</th>
<th>Private</th>
<th>Vulnerable</th>
</tr>
</thead>
<tbody>
<tr>
<td>First name</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Last name</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Email address</td>
<td>Yes</td>
<td>User</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Date of birth</td>
<td>Yes</td>
<td>User</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Sex</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Mobile phone number</td>
<td>No</td>
<td>Contacts</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Current email address</td>
<td>No</td>
<td>Contacts</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Location (country)</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Picture</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Friends</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Yahoo</td>
<td>No</td>
<td>User</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Hotmail</td>
<td>No</td>
<td>User</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Famous Brands/People</td>
<td>No</td>
<td>User</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Job</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Function</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Education</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Graduation Year</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Location</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
</tbody>
</table>

- **Location (country):** This element provides information about the country where the person that registers is living in.

- **Picture:** Same meaning as the Profile picture Facebook data field.

- **Friends:** Same meaning as the Friends Facebook data field.

- **Yahoo:** This element contains the Yahoo identity information of the registering person.

- **Hotmail:** Same meaning as the Windows live id Facebook data field.

- **Famous Brands/People:** This element suggests famous brands and people the person might have an interest in following in.

- **Job:** This element provides the working position of the person that is willing to register.

- **Function:** This element provides the function of the working position of the person that is willing to register.

- **Education:** This element provides the current and historical education the person that is willing to register followed or is following.

- **Graduation Year:** This element provides the year of graduation of the person that is willing to register.

- **Location:** This element provides the location of residence of the person that is willing to register.

From Table 2 can be derived that there are twelve data fields that satisfy the definition of private data fields. Furthermore, according to the definition of vulnerable data, three data fields can be considered as vulnerable: First name, Last name and Sex.

With Google+, the default accessibility setting for a new posting is Contacts, so it will be visible only to other members who are in one or multiple circles of the subscriber.

#### 3.3 MySpace

Similar to the Google+ account, when signing up for a MySpace account, the user has to select whether or not he or she already has an account with Facebook. In case the user has such an account, then the Facebook credentials can be used to sign-up. In this research, it is considered that the user does not have an account with Facebook. This is done in order to use the same sign-up assumptions for all social networking sites.

As mentioned in Section 2.3, Myspace offers the possibility for artists to create special accounts. In this paper it is considered that the user creates an account of type Personal.

An overview and classification of the data fields collected during the Myspace sign-up process can be seen in Table 3.

The data fields that are collected during the Myspace sign-up process and are shown in Table 3 are:

- **First name:** Same meaning as the First name Facebook data field.

- **Last name:** Same meaning as the Last name Facebook data field.

- **Email address:** Same meaning as the Email address Facebook data field.

- **Date of Birth:** Same meaning as the Date of Birth Facebook data field.

- **Sex:** Same meaning as the Sex Facebook data field.

- **Picture:** Same meaning as the Profile picture Facebook data field.

- **Myspace URL:** This element provides a website address with a unique number to link to the Myspace account of the person that is willing to register.

- **Country:** Same meaning as the Email address Google+ data field. Myspace automatically tries to detect and suggest this.

- **Region:** This element provides information about the region where the person that registers is living in.
Table 3. Overview and classification of data fields during Myspace sign-up process

<table>
<thead>
<tr>
<th>Data field</th>
<th>Mandatory</th>
<th>D.A.S.</th>
<th>Private</th>
<th>Vulnerable</th>
</tr>
</thead>
<tbody>
<tr>
<td>First name</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Last name</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Email address</td>
<td>Yes</td>
<td>User</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Date of Birth</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Sex</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Profile picture</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Myspace URL</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Country</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Region</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>City</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
<tr>
<td>Celebrities</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>GMail</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Yahoo</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Rediff</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>IndiaTimes</td>
<td>No</td>
<td>Public</td>
<td>Yes</td>
<td>No</td>
</tr>
<tr>
<td>Friends</td>
<td>No</td>
<td>Public</td>
<td>No</td>
<td>No</td>
</tr>
<tr>
<td>Status</td>
<td>Yes</td>
<td>Public</td>
<td>Yes</td>
<td>Yes</td>
</tr>
</tbody>
</table>

- City: This element provides information about the city where the person that registers is living in. Myspace automatically tries to detect and suggest this.
- Celebrities: Same meaning as the Famous Brands / People Facebook data field, but only for artists.
- GMail: This element contains the GMail identity information of the registering person.
- Yahoo: Same meaning as the Yahoo Google+ data field.
- Rediff: This element contains the Rediff identity information of the registering person.
- IndiaTimes: This element contains the IndiaTimes identity information of the registering person.
- Friends: Same meaning as the Friends Facebook data field.
- Status: This element provides the relationship status of the person that is willing to register. This data field is automatically created and set to ‘single’ when an account is created, without consulting the user.
- Configurability of private data fields: This binary (Yes or No) criterion indicates whether or not the social networking site allows the user to alter the accessibility settings associated with the private data fields, see Tables 1 to 3.
- Number of private data fields: This criterion represents the number of private data fields, see Tables 1 to 3.
- Number of (default) vulnerable data fields: This criterion represents the number of vulnerable data fields, see Tables 1 to 3.
- Supported accessibility settings: This criterion lists the supported accessibility settings (as defined in Section 2) which are supported by this social network.
- Default accessibility setting of postings: This criterion identifies which of the defined accessibility settings is used by default when posting on the social network. This criterion is not associated with the social network sign-up process, but concerns the actual usage of the social network during a posting activity.

From Table 3 can be derived that almost all (fifteen out of seventeen) data fields satisfy the definition of private data fields. Additionally, according to the definition of vulnerable data, eight data fields can be considered as vulnerable: First name, Last name, Date of birth, Sex, Myspace URL, Country, City and Status.

The default accessibility setting for new postings on Myspace is Public.

4. COMPARING SOCIAL NETWORKS BASED ON ACCESSIBILITY SETTINGS

This section provides a comparison of the introduced social networks based on among others, the privacy and vulnerability of the social network data provided in Section 3.

4.1 Criteria

This subsection provides the criteria that are used to compare the social networking sites. These criteria are:

- City
- Celebrities
- GMail
- Yahoo
- Rediff
- IndiaTimes
- Friends
- Status
- Configurability of private data fields
- Number of private data fields
- Number of (default) vulnerable data fields
- Supported accessibility settings
- Default accessibility setting of postings

From Table 3 can be derived that almost all (fifteen out of seventeen) data fields satisfy the definition of private data fields. Additionally, according to the definition of vulnerable data, eight data fields can be considered as vulnerable: First name, Last name, Date of birth, Sex, Myspace URL, Country, City and Status.

The default accessibility setting for new postings on Myspace is Public.

4.2 Comparison

This subsection provides the comparison of the three social networking sites using the criteria discussed in Section 4.1. An overview of the data collected in previous sections is given in Table 4.

The rows in Table 4 correspond to the investigated social networking sites. The columns represent the criteria given in Section 4.1. The element written on the crossing of each row and column shows how that social network scores in terms of the given criterion. What that value of this element represents is given in the definition of the relevant criterion, see Section 4.1.

For each criterion, the scores of the social networking sites will be compared and ranked. This will be done from a privacy and vulnerability perspective, meaning that results which are in favour of privacy and the least vulnerable are considered being the best. The resulting score of all criteria will lead to a ranking of the social networking sites, and thus provide an answer to research question (6).
The first criterion is **Configurability**, which asserts whether or not the social network offers the possibility to alter the accessibility setting of a private data field. Each investigated network scored a ‘Yes’ here, indicating that accessibility settings can be changed at all networks, so they tie for the first spot.

The second criterion is **# Private**, the amount of private data fields the social network requests the registering subscribers to provide. Facebook, Google+, and MySpace request respectively seven, twelve, and fifteen fields, which ranks them 1, 2, and 3, respectively.

**# Vulnerable** is the third criterion, which represents the amount of vulnerable data fields the social networking sites require during the sign-up process. Facebook and Google+ tie for a first place, they both require three fields (**First name**, **Last name** and **Sex**). MySpace ranks as last, requiring a total of eight vulnerable data fields, including the three that Facebook and Google+ require.

The fourth criterion is the **Supported settings**. The possible accessibility settings refer to the ones defined in Section 2. Again, Facebook and Google+ score equal. They support the settings: **Public**, **Contacts**, **User** and **Custom**. MySpace differs from the other two social networks by not supporting the **Custom** setting, while it does support the **Members** setting. This accessibility setting however, is more vulnerable than **Custom**. This is why MySpace ranks last for this criterion, and Facebook and Google+ again tie for the first spot.

The fifth and final criterion is the **Post setting**, which indicates the accessibility setting that is by default applied to the information posted within the social network. For this criterion, Google+ scores as first since its default setting is **Contacts**, which is less vulnerable than **Public**, being the accessibility setting supported by Facebook and MySpace.

From this analysis, it can be concluded that from these three social networking sites, MySpace is the most vulnerable during the sign-up activity and the most vulnerable to subscribers. This is mostly evident from the fact that it has by far the largest amount of private and vulnerable data fields. Even though the accessibility setting of data fields can be configured, it offers more open accessibility settings and more important, the default setting for postings is **Public**.

Facebook and Google+ score equally in three out of the five criteria: configurability, amount of vulnerable data fields and supported accessibility settings. They score differently on the amount of private data fields requested during the sign-up process and on the default accessibility settings for postings.

One of the differences is related to the **amount of private data fields**. It could be considered that Facebook outperforms Google+ from the point of the amount of private data fields, since it requests seven versus the twelve private data fields that are requested by Google+. However, it is worth noting that whilst these data fields yield private data, they are not per se mandatory to fill in.

The second difference between Facebook and Google+ is the default accessibility setting for postings. When a subscriber makes a posting of any sort within the social network, the default accessibility setting is set to **Public** in Facebook and to **Contacts** in Google+. Therefore, Google+ outperforms Facebook in terms of the posting vulnerability.

### 5. CONCLUSIONS AND FUTURE WORK

In this research the accessibility and vulnerability of private data in popular social networking sites was investigated.

Section 2 answered research questions (1) and (2) by analysing and comparing the sign-up process and default posting settings of three of the currently heaviest used social networks. The three social networks that were researched were Facebook, Google+, and MySpace.

Additionally, in this section, the accessibility settings that these social networks offer were classified into five general accessibility settings. The possible settings were: **Public**, **Members**, **Contacts**, **User**, and **Custom**.

Section 3 answered research questions (3), (4) and (5) by defining the private and vulnerable data.

In particular, for each social network, a new account was created, and during this sign-up process it was recorded what data fields were requested. For each data field it was determined whether the data field was mandatory, public available and private. If these three conditions hold, then the data field was considered to be vulnerable.

Additionally, per social network it was recorded what the default accessibility setting for new postings is.

Research question (6) was answered in Section 4, where the gathered data was compared. In order to compare the three social networks, five criteria were defined. These criteria were: configurability of private data fields, number of private data fields, number of vulnerable data fields, supported accessibility settings and default accessibility settings of postings.

The results of this comparison have shown that the three most popular and heaviest used social networks are different in how vulnerable they are with signing-up and making postings. In particular, by analysing how each social networking site performed in terms of the derived criteria, it could be concluded that Facebook is the least vulnerable during signing-up and Google+ is the least vulnerable during postings of information within the social network. MySpace is the most vulnerable regarding both social networking processes.

This research, however, analysed a mere three social networking sites. These sites are the most popular and the heaviest used at this moment, but it could be that other social networks will take over this function in the future. In that case, this research provides a methodology that could be used as a basis for future research.

Other social networking sites could be assessed in the same way as was done in this research. This could provide a means to compare more social networks on the aspects of accessibility and vulnerability of private data.

<table>
<thead>
<tr>
<th>Social Network</th>
<th>Configurability</th>
<th># Private</th>
<th># Vulnerable</th>
<th>Supported settings</th>
<th>Post setting</th>
</tr>
</thead>
<tbody>
<tr>
<td>Facebook</td>
<td>Yes</td>
<td>7</td>
<td>3</td>
<td>Public, Contacts, User and Custom</td>
<td>Public</td>
</tr>
<tr>
<td>Google+</td>
<td>Yes</td>
<td>12</td>
<td>3</td>
<td>Public, Contacts, User and Custom</td>
<td>Contacts</td>
</tr>
<tr>
<td>MySpace</td>
<td>Yes</td>
<td>15</td>
<td>8</td>
<td>Public, Members, Contacts and User</td>
<td>Public</td>
</tr>
</tbody>
</table>

---

**Table 4. Facebook, Google+ and MySpace comparison**
6. ACKNOWLEDGEMENTS
I would like to say a big ‘thank you!’ to my supervisor dr. ir. G. Karagiannis for guiding me through the course for which this paper was written. Thanks for providing a lot of feedback and helping distilling my broad and vague plans into the right scope and concreteness. Also, I would like to thank my fellow students for their reviewing and other input they provided.

7. REFERENCES