Measuring Profile Distance in Online Social Networks

Niklas Lavesson
School of Computing
Blekinge Institute of Technology
SE–371 32, Karlskrona, Sweden
Niklas.Lavesson@bth.se

Henric Johnson
School of Computing
Blekinge Institute of Technology
SE–371 32, Karlskrona, Sweden
Henric.Johnson@bth.se

ABSTRACT

Online Social Networks (OSNs) provide new ways for people to communicate with one another and to share content. OSNs have become quite popular among the general population but their rapid growth has raised concerns about privacy and security. Many predict that the OSNs of today provide a glimpse of the future Internet infrastructure. Whether or not that will be true is difficult to say but what is certain is that the privacy, integrity, and security issues and concerns need to be addressed now. In fact, the mainstream media have uncovered a rising number of potential and occurring problems, including: identity theft, unauthorized sharing of private information, malicious behavior of OSN services and applications, and so on. This paper addresses several important security and privacy issues by focusing on one of the core concepts of OSNs; the user profile, which both includes private and public information that the user shares to different parties and the customized security and privacy settings of the user. We present a method for comparing user profiles, by measuring the distance between the profiles in metric space, and for determining how well an OSN application conforms to user privacy settings. We report on a case study in which the proposed method is applied to Facebook to demonstrate the applicability of the method as well as to motivate its theoretical foundation.

Categories and Subject Descriptors

H.4 [**Information Systems Applications**]: Miscellaneous; I.2.1 [**Artificial Intelligence**]: Applications and Expert Systems—*online social networks, security settings, automatic configuration*

General Terms

Algorithms, Measurements, Security.

Keywords

Profile, application, Euclidean distance.

WIMS'11 May 25-27, 2011 Sogndal, Norway

1. INTRODUCTION

The popularity of Online Social Networks (OSNs) is rising and the habit of designing and implementing online communities is spreading around the world. The web is empowered with a new mode to adopt the real life social relationships into the digital world. Unlike the traditional web, which revolves around information, documents, and web items, the concept of OSNs revolves around individuals, their connections and common interest-based communities. OSNs are useful when it comes to keeping in touch with friends, relatives and colleagues. Millions of users are using OSNs on a daily basis in order to make new contacts, start research collaborations, perform information sharing and even to conduct political campaigns. Some OSNs are used for professional purposes, such as XING and LinkedIn, where it is possible to discover new business connections. Other OSNs are more friendship-oriented and are primarily used for communication, news feeds, entertainment and photo and video sharing. Notable examples of such OSNs are: Facebook, Orkut and MySpace. These networks provide new and interesting ways to communicate, share, and meet on the Internet. However, the growth and use of these networks have raised issues and concerns related to trust, privacy and security. Many predict that the OSN concept to mimic real human communication provides an indication of the future Internet infrastructure. Therefore, it is of great importance to improve the security and privacy mechanisms in OSNs.

With the radical shift on the number of users in the social networks worldwide, there exists an increased threat regarding trust and privacy leaks as compared to the traditional web sites. Trust has been on the research agenda in several disciplines such as computer science, psychology, philosophy and sociology. The research results show that trust is subjective and varies among people depending on what kind of information is disseminated. Moreover, the OSN users are encouraged to share and distribute a variety of personal information, including interests, cultural, religious and social attributes. Most of the users probably believe that this information is only accessible to the social network and maybe their accepted friends. In reality, the set of sources that could get access to the users' sensitive information could include: advertisers, data aggregators, users who are not defined as friends, and external applications. The variety of external, or third party, applications range from games to advanced communication tools, virtual gift agencies, event management systems, and so on.

When analyzing privacy in OSNs, it also apparent that even if the individual user's privacy awareness is high and

several measures are deployed to ensure privacy, the social network could still encounter privacy violations performed by the provider of the social network service. All user data that are directly or indirectly supplied by the user are stored on the provider's servers, which could then possibly be exploited in many ways. The importance of this privacy exposure is the capitalization of the OSN providers such as Facebook or MySpace. The providers' success in attracting a large amount of users who are prepared to share their user profile has increased the market value for each OSN. In order for the providers to continue growing and attracting more users it is important to consider the privacy concerns. One of the major challenges will probably be to find a good trade-off between the level of privacy and the ease of use (user friendliness) of modern OSNs. Thus, the more we can do to aid the individual user with simple, personalized, and automized methods for privacy enhancement the better the situation will become for both OSN providers and their users.

For most regular users, it is often a quite complex venture to adjust the privacy settings to an acceptable level. It is difficult for the users to know about and be able to control the entities that can gain access to their information, i.e., which friendship request should be accepted or which application should be used. For example, when an external application in Facebook is added by a user, the user must decide and grant access to the specific application. The application can then use the private information, i.e., the owner of the application is free to collect, distribute and perhaps misuse the information. The Facebook *Terms of Use* agreement does, however, tell the developer not to misuse information, but it is difficult to put any more restrictions or to be in total control of the development process. To access more information, the applications may ask for subsequent permission grants for each type of information. Therefore, we need better ways to determine how well a user's privacy settings match the permission requests for private data from an external application. It is also the case that the complexity of the privacy controls offered by the providers as well as the trust assumptions of some users make the overall problem difficult to solve.

1.1 Aims and Scope

The aim of this paper is to provide a new context, or perspective, by which user profiles and applications can be analyzed. The idea is to provide a means for mapping the users' privacy settings to the application requests that are carried out to gain access to private user information. Using this context, we present a method for measuring the distance between user profiles and the distance between a certain user profile and a number of applications of interest. In addition, we define personal spheres, which are essentially bounding boxes that enclose a certain user profile and other user profiles and external applications that are based on similar notions of privacy. Our perspective and method are first presented in a general, theoretical context. We then report on a case study in which we apply the method to measure distances between user profiles and applications in one of the major OSNs. This study specifically targets the drawback of today's user-regulated access control methods that are unable to automatically map the profile and policy settings of users to the permission requests of OSN applications.

1.2 Outline

Section 2 sheds some more light on the background of the study, related to privacy issues, malicious activities, terminology, and related work. Section 3 then describes the methodology of how to represent profiles in a metric space and later match them to determine how well an application fits into the personal sphere. Section 4 reports on a case study performed using Facebook, in which the proposed access control methods is analyzed to determine its applicability as well as to motivate the theoretical foundation. Finally, Section 5 features analysis and discussions and Section 6 concludes the paper and gives some pointers to future work.

2. BACKGROUND

OSNs have, by nature, several properties that make them susceptible to exploitation through malicious activities. Some of these properties are: (1) OSNs generally contain a large number of users, (2) the users have developed trust between each others and share similar interests, (3) it is easy for a user to register and create a personal profile, and (4) it is easy to develop new applications for users to accept and be spread among the trusted friends. Adding a friend then involves a confirmation step and the view of the user's profile is then normally limited to Friends Only or to Friends of Friends, unless the user wants the profile to be visible for Everyone on the web.

Facebook is considered to be one of the most popular OSN sites. An interesting feature in Facebook is the ability to create your own applications. Developers with a novel idea and some basic programming skills can quite easily create Facebook applications. Thus, today there exist several hundred thousands of applications [1]. The scope of the applications can be related to entertainment, utilities, productivity or anything you can imagine. If a user wants to install an application, he or she must accept (and thereby grant) that application its requests pertaining to third-party use of private information. The same privacy setting is then used for all applications. Therefore, it is recommended to follow the security *principle of least authority* [18] which states that an application, user or process should only get access to information and resources that are necessary for its legitimate purpose.

Generally we define the security objectives for OSNs in the following categories: privacy, integrity and availability. This work focuses on the privacy issues even though the other two areas are of importance as well but are regarded as out of scope for this study. In accordance to other studies [7] [19] we also believe that the main objective for OSNs is related to privacy and access control, which is the ability for a group or an individual to seclude information about themselves and thereby reveal themselves selectively. The content of what is considered as private differ between individuals and between various cultures. This fact has been discussed by several social network theorists [12] [13], in which the relevance of relations regarding depth and strength have been highlighted. Privacy is in some situations related to anonymity, or the wish to remain unnoticed or unidentified in the public domain. It also calls for the possibility to hide information about the participation in the OSN. Moreover, it is necessary to have accurate default settings that in the first place help the users action to be hidden from any other party (external or internal) to the system. If the user wants to modify

the settings and explicitly disclose any actions this should be made as a second alternative and step. Moreover, as large scale social networks become more common, the amount of information that is stored makes it difficult for an individual to control the level of access that other users should have to his or her information. Similarly, the stored information could potentially be sold by a third party to others and used for purposes not known to the user in question.

The Internet in general and OSNs in particular have brought *new* concerns to privacy since many of our endeavors leave traces and can be permanently stored (for example: blog entries, Twitter posts, status updates and photo tags). The relationship between a person's social network and his or her view on privacy is however a complicated matter. In some situations we want personal information to be spread only to a certain level or to a certain group of people. In other instances, we are prepared to reveal personal information to total strangers. As recently reported, our choices could actually have an effect on our employment since human resource professionals and recruiters are now diligently conducting online research of candidates. This development has created a need for many to control their online privacy settings in addition to controlling their reputation (online as well as offline).

2.1 Terminology

A *user profile*, in the OSN context, is a collection of personal data associated with a specific user. A user profile can store the user's interests, gender, birthday, religious beliefs, and other characteristics of the user. This information can then be exploited by systems, applications or other users in the OSN. In Section 3 we are going to elaborate on the concept of user profiles and we will show of they can be compared and visualized.

An *adversary* is either a user, a malicious OSN provider, a party with access to the infrastructure or a malicious third-party application provider. The adversaries can primarily be seen as a legitimate user but then start to act with bad intentions. For our purposes, the set of adversaries includes, but is not limited to: inside attackers (OSN, user or application provider), and external attackers (not a legitimate participant in the OSN).

Interpersonal ties are defined as connections between people, in which information is carried. Interpersonal ties, is often divided in three varieties: strong, weak, or absent. The strength of an interpersonal tie is a linear combination of the amount of time, the emotional intensity, the intimacy, and the reciprocal services, which characterize each tie [12].

After a user has joined an OSN, they are prompted to identify others in the network with whom they have a relationship. Depending on which OSN the label for these *friendship* differs: Friends, Contacts or Fan. Most OSNs require a bi-directional confirmation of the friendship. The one-direction ties are normally called as Fans or Followers. However, the term Friends can be misleading since the the social connection does in some situation not necessarily mean friendship and the reason why people connect are varied. Therefore, an interesting research area is the development of new methods to classify friendship and how to measure friendship intensity [14] in OSNs.

The *Facebook Platform* provides a set of tools that enable third party developers to integrate with the OSN users. This could be through applications on Facebook.com or external websites and devices. Facebook applications have two core components: a homepage and a profile box. Developers can then choose if the homepage content is proxied through Facebook or isolated in an iframe. For the proxied content the Facebook Markup Language (FBML) is used [9].

2.2 Related Work

A considerable amount of work has been conducted in the privacy and access control field related to traditional Web sites and services. There is a growing amount of work on OSNs as well, however, the focus of such work is often put on the infrastructure and trust between users instead of focusing on the specific issue of how well an untrusted application conforms to the settings defined in a user profile.

An interesting project related to improved OSN security is Safebook [8] in which the main idea is to leverage the trust relationships that are part of the social network itself. Safebook is a decentralized privacy preserving OSN that is governed by the aim of avoiding centralized control over the user data. Decentralization is provided by the use of peer-to-peer technology.

In [15], techniques for building malicious applications in social networks is presented as well as three proof-of-concept examples. The authors do not propose any solutions to the problems but rather show that social networking web sites have the ideal properties to become attack platforms (Anti-social Networks). Possible exploits can then turn OSNs into platforms for malicious and illegal activities, like privacy violation, Distributed Denial-of-Service (DDoS) attacks, disk compromise, malware propagation and personal information leakage. The authors also provide an example of how the adversary can borrow the name of an existing famous application and then create an application with a similar name but with different motives. Through this, the popularity will increase among the users in the social network and the adversary can collect sensitive information to its web server. Notice that, it is only the name of the application that has to be similar not the software itself.

The work presented in [17] describes a new application of threshold-based secret sharing in a distributed OSN. The developed mechanism will select the most reliable delegates based on an effective trust measure. Relationships between the involved friends are used to estimate the trustworthiness of a delegate.

In [9] the privacy risks are addressed associated with social networking Application Programming Interfaces (APIs) by presenting a privacy-by-proxy design. The authors have also performed an interesting study of 150 popular Facebook applications and nearly all applications could maintain their functionality using limited profile information. The results further show that, since the application is given full access to private data, more than 90 % of the applications are given more privileges then they require.

In a study conducted by C. Xi et al. [6], two different forms of private information leaks in social networks are discussed and several protection methods are reviewed. However, most of the current privacy improvement solutions add substantial amounts of user interface complexity or violate social manner. A good interface should not restrict or block users from contributing, sharing or expressing. Thus, a privacy preserving method within social norms is a difficult yet important research aim.

Gilbert and Karahalios [10] reflect upon the fact that so-

cial media treats all users the same: trusted friend or total stranger. In reality, Gilbert and Karahalios argue, relationships fall everywhere along this spectrum and in social science this topic has been investigated for a long time using the concept of tie strength. A quantitative experiment conducted in [10] shows that a predictive model that maps social media data to tie strength using a data set of over 2,000 social media ties manages to distinguish between strong and weak ties with over 85% accuracy. Tie strength, in this context, can be regarded as a means to measure distance in OSNs. Similar approaches have been presented, e.g., to measure friendship intensity [2] and interaction intensity [3] in Facebook. What these studies have in common is that they measure distance on the basis of the observed interaction between users. On the contrary, in the presented study, distance is measured based on user profile data. Essentially, this means that, the former methods require interaction information to measure distance while the method we present can measure distance preventatively between users that have not yet interacted.

Besmer and Lipford examine in users' motivations, intentions, and concerns about using applications, as well as user perceptions of data sharing [4]. The results of this study indicate that the social interaction that stimulates the use of applications also leads to a lack of awareness about data sharing, that is, the risks and possible implications. The study concludes that malicious application that harvest profile information represent a serious threat to user privacy and security and that users are generally not aware of the risk of information theft. Besmer and Lipford suggest that a potential solution to this problem is to improve the integration of privacy settings into the graphical user interface of the most common interaction tasks. Besmer et al. further present a user study of an interface design for setting OSN privacy settings [5]. The results indicate that the proposed model and interface are suitable for users who are concerned about their privacy and are aware of different privacy issues. However, it is also stated that more work is needed to explore alternate means for creating policies for users with a greater lack of privacy awareness.

To improve what can be perceived as the current infrastructure-based defects of OSN, the Davis Social Link (DSL) research group[1] has developed an infrastructure, which is referred to as FAITH (Facebook Applications: Identifications Transformation & Hypervisor). The objective of FAITH is to provide more trustworthy and flexible services to users. FAITH itself is technically a Facebook application, but it acts as a proxy between Facebook users and the Facebook infrastructure. It can hook other applications and provide services to these applications. FAITH provides two major functions related to the presented study:

1. Each application hooked under FAITH is monitored by FAITH. All Facebook APIs that the application called are logged and available for the user to review. The log information helps the users to keep track of their personal information executed by an application and also for the system to perform anomaly/intrusion detection.

2. Users can customize the API related to their personal information that an application can call. If a user feels an API is not necessary for an application, he/she can block the API so that the application can not access the user's relevant information. API customization prevents applications from maliciously impersonating the user.

[1]http://dsl.cs.ucdavis.edu/

In conclusion and to the best of our knowledge, we have not been able to find any work that focus directly on the methodology of comparing (user and application) profiles for determining or visualizing how well the profiles conform to certain privacy settings. In fact, related approaches for computing distances between OSN users (e.g., the strength of their friendship) are most often based on information about the interaction between the users.

3. METHOD

We will now introduce a new perspective on OSN user profiles and present a method for comparing user profiles and for determining whether or not an OSN application conforms to the privacy settings of a particular user. First, we describe how user profiles can be defined in a metric space, which enables user profile comparison. Second, we show how this profile comparison can be conducted by measuring the distance between user profiles in the metric space. Third, we introduce the concept of personal spheres within the context of OSN user profiles and describe how OSN application permission requests can be transformed to the metric user profile space and subsequently compared with a user's user profile and his or her personal sphere.

To summarize, we present methods for measuring profile-to-profile distance or profile-to-application distance. The former can be used, for example, to determine which OSN friends to share media and information with and the latter can be used to determine whether or not the permission requirements of an OSN application conform to the privacy settings of one's user profile.

3.1 Representing Profiles in Euclidean Space

An OSN user profile can be quite complex and can, for example, include: basic personal information, contact information, marital status, date of birth, educational history, but it may also include or be associated with: the list of OSN friends, the list of blocked applications and users, or any other type of information or personalized setting. Ultimately, an OSN user profile is essentially just a list of values, where each value is associated with a parameter. On a higher level of abstraction, of course, sets of parameters can be associated with different parameter families or groups. Consider, for example, that parameters could be organized into the following two groups: user data and privacy settings. Elaborating on this example, each parameter group could be further divided into sub groups out of which the first example group of ours could include the following sub groups: image data, text data, video data, and sound data.

Let P represent the user profile space. A user profile can be defined as a tuple, $p \in P$, of n elements, $e_1, \ldots, e_n$, where each element, e_i, represent a profile parameter setting, i. For our purposes, it is sufficient that elements of p can represent either discrete or continuous variables, meaning that the profile could include information such as: date of birth (continuous), gender (discrete), general privacy level (discrete) and so on.

3.2 Profile Distance

By using the aforementioned profile definition, any profile, $p_j \in P$, can be represented by a point, j, in the real coordinate system, more specifically, the n-dimensional Euclidean space. Moreover, the distance between two profiles, p_j and p_k, can thus be calculated using the Euclidean distance metric:

$$d(j,k) = \sqrt{(j_1 - k_1)^2 + \cdots + (j_n - k_n)^2} = \sqrt{\sum_{i=1}^{i<n} (j_i - k_i)^2}$$

It is possible to enhance the usability of profile distance measurement by introducing a concept we like to denote the personal sphere: conceptually, we could describe this sphere as a combination of pieces of knowledge, experiences, psychological perspectives, preferences, and so on, that collectively describe an individual or a group of similar individuals. Add to this, that individuals outside of one's personal sphere may be regarded as different or dissimilar to one's own character. In fact, the sphere could define anything from a stratum of society to a particular field of activity. However, for our purposes, the personal sphere could be seen as enclosing similar user profiles. Similar, in this context, refers to the notion that the two user profiles share essential features, e.g., with respect to privacy settings or age group.

3.3 The Personal Sphere

Using previous definitions, we are able to define the personal sphere as an ordinary sphere (or a hypersphere if $n > 3$) in the mathematical sense, using a central point that is equal to the profile point in question, say p_j, and a reasonable radius, r. We say that a user profile, p_k, is similar to p_j if it satisfies the following inequality and thus is included in the personal sphere of p_j:

$$r \leq \sqrt{(k_1 - j_1)^2 + \cdots + (k_n - j_n)^2}$$

We provide a simple example visualization of user profile distances and personal spheres in Figure 1. For cases where $n > 3$, it is also possible, and perhaps more feasible, to make use of multiple radii and a more general bounding box instead of a bounding sphere. Alternatively, the n-dimensional space can be transformed into a two-dimensional or three-dimensional space using multi-dimensional scaling in order to visualize the personal sphere.

3.4 User Profiles and Applications

The aim of this paper, as mentioned earlier, is two-fold: we provide a way to measure the distance between different user profiles but, perhaps more importantly, we address the question of how to quantify how well an OSN application conforms to the personal sphere of an OSN user. This is done by establishing a way to measure the distance between a user profile and a certain application.

In the general case, an OSN service or an OSN application gives the user a certain functionality but more often than not, the application provides this functionality with some strings attached. A common argument is that, in order to personalize the service provided and to increase the social networking experience, the application may ask for certain permissions that can, for example, be associated with the extent to which private profile data is shared or whether or not the application can act on the user's behalf for certain actions.

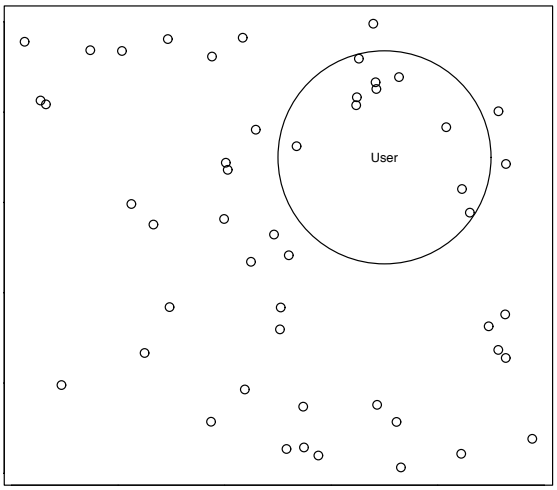

Figure 1: A two-dimensional visualization, obtained using multi-dimensional scaling, of a personal sphere around a particular user depicted together with multiple friends (small circles).

For a majority of the large-scale OSNs available today, the permissions are directly related to settings that are customizable within the realms of the user profile. Thus, not only does the user need to customize general privacy and security settings in the user profile, the user also needs to consider the very same privacy and security aspects each time an application is to be used for the first time (and sometimes even on multiple occasions during the use of an application). In the OSNs reviewed, there is not a clear one-to-one mapping between application permission queries and user profile security and privacy settings, which means that a user can opt for a certain OSN behavior or privacy level in the user profile while unknowingly permitting conflicting behavior from an application.

In order to address this issue, we therefore let Q represent the application permission query space. Analogously to user profiles, an application permission query can be defined as a tuple, $q \in Q$, of m elements, $f_1, \ldots, f_m$, where each element, f_l, represent a required application permission, l. Again, for our purposes, it is sufficient that elements of q represent discrete variables. In fact, a Boolean variable is usually sufficient since a required application permission can be represented by a value of 1 whereas 0 refers to a permission that is not required. In order to measure the distance between the required permissions of an application and the personal sphere of a user, we need to provide a mapping (a function) between the user profile space, P, and the application permission query space, Q:

$$g : Q \rightarrow P$$

The function, g, can be quite complex and will vary substantially between different OSNs. However, theoretically, we may regard g as a general purpose mapping and it can thus be used to measure the distance between a user profile and an application or to find out whether an application is within the personal sphere of a user: an application permission query, $q_k \in Q$ is first mapped to P which yields

Table 1: Sharing settings excerpted from Facebook with example choices, marked with X, for each setting.

Setting	Everyone	Friends of Friends	Friends Only	Other
My status, photos, and posts				X
Bio and favorite quotations			X	
Family and relationships				X
Photos and videos I'm tagged in			X	
Religious and political views			X	
Birthday	X			
Can comment on posts			X	
Places I check in to			X	
Contact information		X		

$p_k \in P$. The query can now be regarded as a user profile and may thus be compared with personal spheres or other user profiles by way of distance measurement.

We now present a case study in which we show how to apply the proposed method in a real-world OSN setting. The case study not only demonstrates the practical use of our method but also serves as a way to motivate and explain the theoretical basis of the method.

4. CASE STUDY

Our study focus on Facebook, as it is currently the most widely used application platform. Facebook reports having 500 million active users (users who have returned to the site in the last 30 days)[2]. Due to the popularity of Facebook we decided to perform a case study on it using the proposed methodology. As described below, Facebook gives the users the opportunity to set their privacy setting to be locked and visible only to their friends. However, a Facebook application can with or without the permission from the user collect private information by using the calling methods of Facebook API. Previous study [9] has shown that although users are able to set the privacy setting each time an application is installed most users give all the applications full access to their accounts. Therefore, an adversary can collect sensitive information and store it on an external web server collaborating with the malicious application.

4.1 Facebook Sharing

The following section describes Facebook's privacy policy and practices[3]. When a user runs an application, personal information (including profile information) is available to the application developers. Facebook requires the developers to agree to terms that limit their use of the information and further claim that they use technical measures to ensure that they only obtain authorized information. When a user connects with an application or website it will have access to general information. The term general information includes the name of the user and its friends' names, profile pictures, gender, user IDs, connections, and any content shared using the Everyone privacy setting. Facebook may also make information about the location of a user's computer or access device and age available to applications and websites. If the application wants to access any other data, it will have to ask for the user's permission.

Facebook recommends the users to review the policies of third party applications to make sure that the users are comfortable with the ways in which they use the information shared with them. This is however not a trivial task to perform depending on the user's knowledge of how to adjust the settings or if the application seems to be trusted because other friends are using it. Therefore, if a friend connects with an application, the application will then be able to access the name, profile picture, gender, user ID, and information shared with everyone. It will also be able to access connections, except it will not be able to access the list of the user's friends. If the user has already connected with (or has a separate account with) that application, it may also be able to connect its friends on that application.

Facebook further states that they provide a safe, efficient, and customized experience. In order to obtain this they serve personalized advertising, allow advertisers to choose the characteristics of users who will see their advertisements and they may use any of the non-personally identifiable attributes collected (including information you may have decided not to show to other users, such as your birth year or other sensitive personal information or preferences) to select the appropriate audience for those advertisements. Facebook further states that they occasionally pair advertisements with relevant information about users and their friends to make advertisements more interesting and more tailored. It is also possible that they use information collected from other Facebook users to supplement a profile.

Certain downloadable software applications and applets that Facebook offers, such as browser toolbars and photo uploaders, transmit data to Facebook. The policy further states that they may not make a formal disclosure if they believe that the collection of and use of the information is the obvious purpose of the application.

4.2 User Privacy Settings

Facebook provides user-customizable settings that govern which type of profile information is shared and to whom. Table 1 presents the nine settings available today. There are four possible choices for each setting. Sharing can be allowed for: Everyone, Friends of Friends, or Friends Only. An additional choice, referred to as Other, can be used to make a more detailed decision about sharing. Facebook includes functionality for creating groups and for including selected users into these groups, referred to as friendlists. The choice of Other can be used, for example, to specify that sharing is allowed to only to friends featured in a specific user-defined group.

[2]http://www.facebook.com/press/info.php?factsheet

[3]Facebook Policy, `http://www.facebook.com/policy.php`

4.3 Information Shared through Friends

The Facebook user can control which information is available to Facebook applications, games and websites when they are used by friends of the user. As can be seen in Table 2, there are currently 18 settings that can be checked or unchecked to provide or restrict this type of access of different pieces of information.

It is worth noticing that a user's name, profile picture, gender, networks and the user's ID (together with any information a user has chosen to set to Everyone) is always available to friends' applications unless the user chooses to turn off platform applications and websites.

Table 2: Settings for which information is accessible through a user's friends, excerpted from Facebook.

Bio	My videos
Birthday	My links
Family and relationships	My notes
Interested in and looking for	Photos and videos I'm tagged in
Religious and political views	Hometown
My website	Current city
If I'm online	Education and work
My status updates	Activities, interests, things I like
My photos	Places I check in to

4.4 Application Permissions

By default, a Facebook application can access all public user profile data, including: name, profile picture, gender, and the list of friends. If the application needs to access other profile data or if the application needs to act on a user's behalf on Facebook, e.g., to publish content on the wall or to perform activities when the user is offline, the application must request extended permissions. Table 3 includes a subset of these permissions. However, the complete set of permissions is available, together with textual descriptions, for Facebook users from an online repository[4].

5. ANALYSIS AND DISCUSSION

We now take what we know from Section 4 in order to apply our method (defined in Section 3) to a practical example. For increased interpretability, we choose to focus solely on Facebook sharing settings (described in Table 1). It would of course be straight-forward to include, for example, the settings for which information is shared through friends (described in Table 2) as well. Table 4 describes 24 user profiles, in which the letters a to i defines the sharing settings in Table 2, i.e., a = My status, photos and posts, and b = Bio and favorite quotation, and so on. We have included a set of standard user profiles from Facebook, namely: Everyone, Friends of Friends, Friends Only, and Recommended. Additionally, we have randomly generated 20 user profiles. Each sharing setting is set to a particular value from the ordinal scale: Only Me (0), Friends Only (1), Friends of Friends (2), and Everyone (3). As is clearly shown, three of the standard user profiles contain the same value for each setting while the Recommended user profile can be described as having less restrictive sharing settings for basic information and more restrictive settings for private information, e.g., regarding contact options and birthday date.

[4]Facebook Extended Permissions, `http://developers.facebook.com/docs/authentication/permissions`

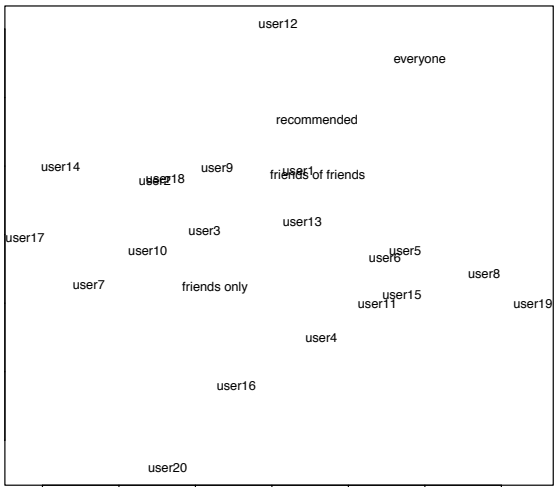

Figure 2: A two-dimensional visualization, obtained using multi-dimensional scaling, of the sharing settings of 24 user profiles: everyone, friends only, friends of friends, recommended, and 20 randomly generated profiles.

We will soon find out how this simple data set can be used to elaborate on how our proposed method works in practice. First and foremost, however, we turn our attention towards the concept of multi-dimensional scaling and its application to user profile distance analysis.

5.1 Scaled versus Unscaled Data

The Euclidean distance between user profiles can either be computed on raw user profile data (for example, from Table 4) or on a two-dimensional user profile data representation obtained through multi-dimensional scaling (MDS), which is basically the family name for a set of statistical techniques that can be used for visualization, especially when the goal is to explore similarities or dissimilarities in data. The initial step in most of these MDS approaches is to compute a matrix of object-to-object (dis)similarities (where objects can be defined as vectors of any dimension). A classical and commonly used approach for MDS then is principal coordinates analysis (sometimes referred to as Torgerson Scaling or Torgerson–Gower scaling) [11], which takes the input matrix (with dissimilarities between pairs of objects) and outputs a coordinate matrix. The configuration of this coordinate matrix minimizes a loss function, called the strain. Given a set, $E \subset P$ of user profiles, a preferred number of dimensions, c, to scale to, and a distance function (Euclidean distance in our case):

$$d(j,k) = \sqrt{(j_1 - k_1)^2 + \cdots + (j_n - k_n)^2}$$

which denotes the distance between user profiles j and k in

Table 3: A subset of the Extended Permissions, together with the associated textual description, excerpted from Facebook. The complete set features 52 permissions, divided into the three groups: Publishing, Data, and Page permissions.

Permission	Description
create_event	Enables your application to create and modify events on the user's behalf
rsvp_event	Enables your application to RSVP to events on the user's behalf
user_about_me friends_about_me	Provides access to the About Me section of the profile in the about property
user_activities friends_activities	Provides access to the user's list of activities as the activities connection
user_birthday friends_birthday	Provides access to the birthday with year as the birthday_date property
user_education_history friends_education_history	Provides access to education history as the education property

Table 4: Sharing settings for 24 user profiles: everyone, friends only, friends of friends, recommended, and 20 randomly generated profiles.

	Sharing setting								
User profile	a	b	c	d	e	f	g	h	i
everyone	3	3	3	3	3	3	3	3	3
friends only	1	1	1	1	1	1	1	1	1
friends of friends	2	2	2	2	2	2	2	2	2
recommended	3	3	3	2	2	2	1	1	1
user1	3	2	2	2	3	0	0	1	2
user2	1	1	1	0	3	1	1	1	2
user3	0	1	2	1	2	0	3	2	3
user4	1	3	0	3	0	1	2	3	1
user5	1	1	1	3	2	1	0	3	3
user6	2	3	1	3	0	3	0	2	0
user7	0	2	1	1	1	1	3	0	1
user8	3	1	3	3	0	1	2	3	2
user9	0	2	3	0	1	2	1	2	0
user10	1	2	0	2	3	1	3	1	2
user11	3	0	0	2	2	1	1	3	2
user12	2	3	3	0	2	2	0	1	3
user13	1	3	2	2	0	1	3	2	3
user14	0	3	1	0	2	1	2	0	2
user15	0	0	3	2	0	0	1	3	3
user16	3	0	1	3	2	0	2	0	0
user17	1	2	0	0	3	0	2	0	1
user18	2	3	0	1	3	2	3	2	1
user19	3	0	1	3	0	2	0	3	3
user20	1	1	0	3	0	0	3	0	0

Columns a to i correspond to the sharing settings in Section 5.2

user profile space, we construct a dissimilarity matrix:

$$\Delta = \begin{pmatrix} d(1,1) & d(1,2) & \cdots & d(1,|E|) \\ d(2,1) & d(2,2) & \cdots & d(2,|E|) \\ \vdots & \vdots & & \vdots \\ d(|E|,1) & d(|E|,2) & \cdots & d(|E|,|E|) \end{pmatrix}$$

Given Δ, our goal is now to find $|E|$ vectors, $z_1, \ldots, z_{|E|} \in \mathbb{R}^c$, such that $d(x_j, x_k) \approx d(j,k)$ for all $j,k \in E$. Our objective, more informally stated, is to find a two-dimensional representation of the high-dimensional user profiles while preserving distances.

Such a two-dimensional representation of the data from Table 4, computed using the principal coordinates analysis multi-dimensional scaling approach, is depicted in Figure 2. The form of representation is attractive in the given context since distances of particular interest can be depicted together with personal spheres in graphs (or in three-dimensional plots if one more dimension is added). However, for this approach to have any practical use the resulting distances must be correlated with the original distances. This is due to that the latter could be regarded as true distances between user profiles insofar Euclidean distance can be regarded as a good approximation of true distance. When we perform a correlation test on the distance matrices generated by unscaled and scaled user profile data, using Pearson's correlation coefficient (R), we declare that, with 95% confidence, the true R is included in the interval $[0.613, 0.741]$. Thus, the correlation between unscaled and scaled user profile data is statistically significant at $Pr < 0.05$ which arguably means that we would be able to confidently compute user profile distances on scaled two-dimensional data without a too great loss of information (or without too much distortion). However, it is generally known that, for MDS cases with a high demand on accurate preservation of distances, it is recommended that R^2 should be at least higher than 0.5 and preferably over 0.9.

We now leave the discussion about multi-dimensional scaling and distance visualization in order to turn to the primary aim of the paper: the measurement of the distance between user profiles and the measurement of the distance between a certain user profile and different applications. As a proof of concept, we are going to measure the distance between a selected set of user profiles and a set of popular Facebook applications.

5.2 The Distance to Top Applications

A number of Internet sites publish trend analyses of OSN applications in general and Facebook applications in particular. One such site maintains a list of the 15 top Facebook applications (related to the number of users). A recently updated list from this site[5] includes the specific applications presented in Table 5 along with the extended permissions they require. All applications require permission to be granted for obtaining basic information about the user (e.g., name, list of friends, and so on). Notably, one particular application stands out since it requires 11 permissions to be granted. In order to compute the distance between these

[5]AppData, `http://www.appdata.com`

Table 5: A Top 15 List of Applications

Application	Permissions
Farmville	user_about
Phrases	user_about
Texas HoldEm Poker	user_about
Frontierville	user_about
Mafia Wars Game	user_about, email
Causes	user_about, email, publish_stream, user_birthday
Cafe World	user_about, email, publish_stream, user_birthday
Quiz Planet	user_about
Give Hearts	user_about, publish_stream
Are You Interested?	user_about, email, user_birthday, friends_online_presence, user_photos, user_videos, user_activities, user_interests, user_hometown, user_religion_politics, user_likes
Treasure Isle	user_about, email
iHearts	user_about, publish_stream
Millionaire City	user_about
Pet Society	user_about, email
Windows Live Messenger	user_about, publish_stream, read_stream

applications and different user profiles, we must first transform the extended permissions required by the applications to user profile vectors. In other words, we must define g in an appropriate way, given the context and possible restrictions and requirements.

5.2.1 Application Profiles

For the purpose of this case study, we argue that it is sufficient to map the permission requests to the sharing settings of a Facebook user profile. After having reviewed the descriptions of each extended permission, we construct the following associations:

a) My status, photos, and posts
- user_videos
- user_status
- user_photos

b) Bio and favorite quotations
- user_about_me

c) Family and relationships
- user_relationships

d) Photos and videos I'm tagged in
- user_photo_video_tags

e) Religious and political views
- user_religion_politics

f) Birthday
- user_birthday

g) Can comment on posts

h) Places I check in to
- user_checkins
- user_location

i) Contact information
- email
- user_hometown
- user_website

That is, given that α represent a particular choice of sharing settings, we define the application user profile as a vector, $p^{\alpha} = \{\alpha_a, \alpha_b, \alpha_c, \alpha_d, \alpha_e, \alpha_f, \alpha_g, \alpha_h, \alpha_i\}$. In this study, we make the assumption that, if a user grants an application an extended permission (e.g., user_videos), the user has decided that the associated information (e.g., My status, photos, and posts) can be shared to everyone. Thus, using the same ordinal scale as in Table 4, we assign 3 (everyone) to the application user profile element in question. For example, if an application requires user_videos, we let $\alpha_a = 3$. Furthermore, we let all unassigned elements take the same values as the corresponding elements in the user profile we want to compare the application with. Thus, if we have Facebook's Recommended user profile, $a = \{3, 3, 3, 2, 2, 2, 1, 1, 1\}$ and wish to measure its distance to the Pet Society application with a raw profile of $b = \{?, 3, ?, ?, ?, ?, ?, ?, 3\}$, the subsequent application user profile would be defined as: $b^{a} = \{3, 3, 3, 2, 2, 2, 1, 1, 3\}$.

5.2.2 Application Cluster Identification

We define an application cluster to be a point in (Euclidean) space at which more than one application user profile reside. By observing the applications and their associated permission requests presented in Table 5, it is evident that three clusters of applications can be identified. Cluster 1 contains the applications: Farmville, Phrases, Texas HoldEm Poker, Frontierville, Quiz Planet, and Millionaire City. Similarly, Cluster 2 contains: Mafia Wars Game, Treasure Island, and Pet Society. Finally, Cluster 3 contains the two applications: Causes and Cafe World. Apparently, the Are You Interested? application is situated on its own unique point in space, that is, it is not located in any of the aforementioned clusters.

5.2.3 Comparison with the Recommended Profile

Using a reasonable personal sphere radius, say $r = 3$, we may now measure the distance between some user profiles of choice and the applications. Which sphere radius is reasonable in practice is of course a matter that should be decided by either a domain expert or the informed user. As an alternative approach, we suggest that a reasonable radius can also be automatically determined by employing supervised learning, clustering or optimization algorithms to retrospective or prospective data depending on the specific problem at hand. For example, it would be possible to randomly select a sample of users, then extract their user profile data and the user profile data of their friends. Next, clustering analysis is performed to determine if any natural friend clusters can be found for each sampled user. If there is correlation between the sampled users (in terms of distances from each user to his or her friend clusters), we could argue that we have established radii that represent different depths or levels of friendship. Another example would be to initialize the sphere radius, again, to a reasonable value and then let the social network ask the user from time to time whether or not it is correctly assumed that a certain (randomly selected) friend reside within the personal sphere. For this semi-automatic, prospective-data based approach, an online learning algorithm could be applied to tune the sphere radius parameter over time.

The Facebook recommended user profile (in terms of sharing settings) is located at acceptable distances to all applications: 0.00 (Cluster 1), 2.11 (Cluster 2), 2.36 (Cluster 3), and 2.58 (Are You Interested?), as is shown graphically in Figure 3. Quite intuitively, the distances to more demanding applications (in terms of permission requests) are longer than the distances to less demanding applications. In a real-world setting, the personal sphere radius could be iteratively and semi-automatically reduced or enlarged by employing a well-known anti-spam software technique: the user could be given the opportunity to classify applications residing inside their personal sphere as privacy-intrusive or vice versa. After this classification, the OSN could change the radius in the right direction (either reduction or enlargement). Of course, such changes to the personal sphere could, and probably would, have other implications on the privacy level of the user.

5.2.4 Comparison with the Friends Only Profile

By using the equivalent personal sphere radius as in the former comparison, we may also compute the distances between a more restrictive user profile, the Friends Only profile, and the same set of applications. The Friends Only user profile then, is located at acceptable distances only to Cluster 1 (2.11) and Cluster 2 (2.98). However, the distances to Cluster 3 (3.65) and the Are You Interested? application (4.71) are too long. What this means is that the Are You Interested? application and Cluster 3 reside outside the personal sphere of the Friends Only user, as can be clearly viewed in Figure 3. The ratio, of the distances computed on the unscaled profiles and the distances computed on the MDS-transformed profiles, is in proximity of 0.99. In other words, the distance information is quite accurately preserved through multi-dimensional scaling.

An interesting point of discussion to be made here is that, the knowledge that can be acquired from analyzing the distance between a restrictive user profile and different applications can be of use to application vendors: as people become more aware of the importance of online privacy and security, vendors may tune the functionality and permission requirements of their applications, using the results from such analyses, to cater to an security-aware audience. If several vendors provide applications with similar functionality but which reside on quite different distances from the potential user's profile, the vendor with the least demanding application may that fact as a selling point.

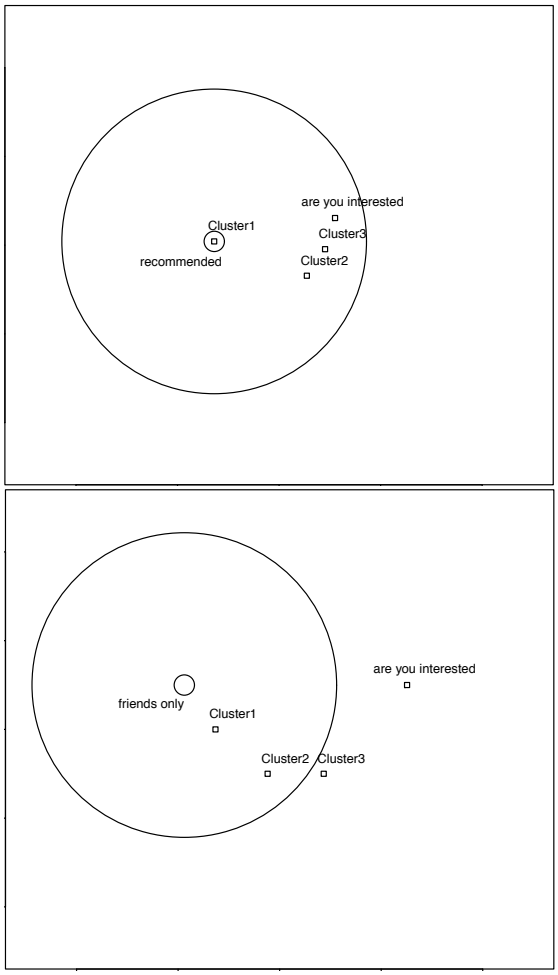

Figure 3: Two-dimensional visualizations, obtained using multi-dimensional scaling featuring comparisons between the surveyed applications and the Recommended and Friends Only user profiles, respectively. A personal sphere with a radius of 3 is used in both plots.

5.2.5 Discussion

The simple examples provided above can easily be extended to include more comprehensive user profile descriptions and this extension can be made by just adding elements to the user profile vector. Optionally, or rather, depending on the context, the application-to-user-profile mapping function, g, can be re-defined more elaborately. No matter how many dimensions the defined user profile vector may have, the distance measurements and the introduction of the per-

sonal sphere will be conducted in the same manner. Thus, essentially, our method is highly generalizable within the Facebook context but arguably it is also generalizable to many of the other major OSN platforms. The perspective and method we have introduced can be used both for visualization and measurements pertaining to user profiles and applications in OSNs. Moreover, it provides the basis for automatically determining whether one or more applications conform to the privacy and security settings of a particular OSN user profile. For example, machine learning or general artificial intelligence techniques can be used to automatically analyze OSN populations in order to answer questions about the extent of which users tend to use applications that reside outside of their personal sphere. Of course, and even more interesting, is the fact that such techniques can be employed to warn users about applications that do not conform to their view of privacy.

Although this paper focuses on privacy and integrity issues and tries to present a method for enhancing user privacy and security, there are many other domains in which our method can be applied. For example, in online dating or job finder social networks. The main functionality of such networks is to provide their users with good matches, that is, profiles that match their own profile. In the context of online dating, recent work has focused on the problem of learning user preferences [16]. Interestingly, this work also states that the problem of inaccurate explicit user preferences is not confined to online dating but rather it is a problem for all domains in which users do not know precisely what they want or are unable to accurately specify their preferences. This very notion is unmistakably central in the presented study as well.

6. SUMMARY

This paper has addressed a number of the security and privacy issues of modern Online Social Networks (OSNs) by focusing on one of their core concepts; the user profile, which usually includes both the private and public information that the user shares to different parties and the customized security and privacy settings chosen by the user. We have introduced a new perspective on user profile and privacy settings analysis, namely by comparing user profiles and applications in the Euclidean space, and we have also presented a method for comparing user profiles, by measuring the (Euclidean) distance between the profiles, and for determining how well an OSN application conforms to the privacy settings defined in a user profile. The latter is carried out by defining a bounding box (denoted in the presented study as the personal sphere) around the user profile and then transforming application permission requests to user profile space in order to determine which applications reside outside the personal sphere of a user.

We report on a case study in which the proposed framework and method are applied to Facebook in order to demonstrate the applicability of our approach as well as to motivate the theoretical foundation. In the context of this case study, we describe how to define user profiles by extracting information about privacy settings.

6.1 Conclusions

The analyses indicate that the proposed framework and method are feasible to use, at least in the Facebook context, and that they can provide users (and application vendors) with a detailed, yet comprehensive, description of which applications may impose on the privacy or integrity of different users. The analysis performed in context of the case study indicates that multi-dimensional scaling is a suitable approach for transforming the user profile comparison results into a comprehensible graphical representation; a profile-distance graph. Based on a theoretical analysis of our method, we conclude that it can be adopted quite easily for use with, for example, techniques from artificial intelligence or machine learning in order to automatically determine a reasonable personal sphere radius and also to provide means for elevating personal privacy and security by automatically warning users about privacy-intrusive applications. However, the method proposed in this paper could also be employed for radically different reasons, for example, to assess user profile similarity in match making applications.

6.2 Pointers to Future Work

For future work, we are going to conduct an extensive survey to collect (anonymized) user profile settings as well as information about application usage statistics, from a large sample of Facebook users. These data would be collected in order to make a real-world assessment of the ratio of users that may be in jeopardy of using applications that are intrusive to their privacy and integrity. For this investigation, we are going to include as much information from the user profile as is computationally feasible to be able to preserve the integrity of the complete user profile when transforming it to Euclidean space. Moreover, in subsequent studies, we aim to prove that our findings are generalizable even to other major OSNs.

7. ACKNOWLEDGMENTS

The authors would like to thank the members of the distributed and intelligent systems laboratory (DISL) at Blekinge institute of technology for their valuable comments and suggestions.

8. REFERENCES

[1] Facebook Analytics and Advertising. http://adonomics.com.

[2] W. Ahmad, A. Riaz, H. Johnson, and N. Lavesson. Predicting friendship intensity in online social networks. In *21st International Tyrrhenian Workshop on Digital Communications*, 2010.

[3] L. Banks and S. F. Wu. All friends are not created equal: An interaction intensity based approach to privacy in online social networks. In *International Conference on Computational Science and Engineering*, 2009.

[4] A. Besmer and H. R. Lipford. Users' (mis)conceptions of social applications. In *Proceedings of Graphics Interface 2010*, GI '10, pages 63–70. Canadian Information Processing Society, 2010.

[5] A. Besmer, H. R. Lipford, M. Shehab, and G. Cheek. Social applications: exploring a more secure framework. In *Proceedings of the 5th Symposium on Usable Privacy and Security*, SOUPS '09, pages 2:1–2:10, New York, NY, USA, 2009. ACM.

[6] X. Chen and S. Shi. A literature review of privacy research on social network sites. In *Proceedings of the 2009 International Conference on Multimedia*

Information Networking and Security - Volume 01, MINES '09, pages 93–97, Washington, DC, USA, 2009. IEEE Computer Society.

[7] G. Crescenzo and R. J. Lipton. Social network privacy via evolving access control. In *Proceedings of the 4th International Conference on Wireless Algorithms, Systems, and Applications*, WASA '09, pages 551–560, Berlin, Heidelberg, 2009. Springer-Verlag.

[8] L. A. Cutillo, R. Molva, and T. Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. *IEEE COmmunication Magazine*, pages 94–101, 2009.

[9] A. Felt and D. Evans. Privacy protection for social networking apis. In *In proceedings of WEB 2.0 Security and Privacy 2008 (W2SP 2008)*, 2008.

[10] E. Gilbert and K. Karahalios. Predicting tie strength with social media. In *Proceedings of the 27th international conference on Human factors in computing systems*, CHI '09, pages 211–220, New York, NY, USA, 2009. ACM.

[11] J. C. Gower. Some distance properties of latent root and vector methods used in multivariate analysis. *Biometrika*, 53:325–328, 1966.

[12] M. Granowetter. The strength of weak ties. *American Journal of Sociology*, 78:1360–1380, 1973.

[13] M. Granowetter. A network theory revisited. *Sociological Theory*, 1:201–233, 1983.

[14] H. Johnson, N. Lavesson, H. Zhao, and S. F. Wu. On the concept of trust in online social networks. *Springer Book - Trustworthy Internet, In Press*, March 2011.

[15] A. Makridakis, E. Athanasopoulos, S. Antonatos, D. Antoniades, S. Ioannidis, and E. P. Markatos. Designing malicious applications in social networks. In *In IEEE Network Special Issue on Online Social Networks*, 2010.

[16] L. Pizzato, T. Chung, T. Rej, I. Koprinska, K. Yacef, and J. Kay. Learning user preferences in online dating. In *Preference Learning Workshop, European Conference on Machine Learning and Principles and Practice of Knowledge Discovery in Databases (ECML/PKDD 2010)*, 2010.

[17] L. H. Vu, K. Aberer, S. Buchegger, and A. Datta. Enabling secure secret sharing in distributed online social networks. In *Proceedings of the 2009 Annual Computer Security Applications Conference ACSAC '09*, pages 419–428. IEEE Computer Society, 2009.

[18] J. Whittaker. Why secure applications are difficult to write. *IEEE Security & Privacy*, 1(2):81–83, Mars 2003.

[19] C. Zhang, Z. Jinyuan, and Y. Fang. Privacy and security for online social networks: challenges and opportunities. *Network. Mag. of Global Internetworking.*, 24:13–18, July 2010.