Security and Privacy in Online Social Networks

Leucio Antonio Cutillo[1], Mark Manulis[2], and Thorsten Strufe[2]

[1] Eurecom, Sophia Antipolis, France,
cutillo@eurecom.fr
[2] TU Darmstadt & CASED, Darmstadt, Germany
mark@manulis.eu
strufe@cs.tu-darmstadt.de

1 Introduction

Social Network Services (SNS) are currently drastically revolutionizing the way people interact, thus becoming *de facto* a predominant service on the web, today[1] The impact of this paradigm change on socioeconomic and technical aspects of collaboration and interaction is comparable to that caused by the deployment of World Wide Web in the 1990's.

Catering for a broad range of users of all ages, and a vast difference in social, educational, and national background, SNS allow even users with limited technical skills to publish ***Personally Identifiable Information*** (PII) and to communicate with an extreme ease, sharing interests and activities.

An ***Online Social Network*** (OSN) offering, usually centralized, online accessible SNS contain digital representations of a subset of the relations that their users, both registered persons and institutions, entertain in the physical world. Spanning all participants through their relationships, they model the social network as a graph. Every OSN user can typically create his or her own *OSN profile* and use the available *OSN applications* to easily share information with other, possibly selected, users for either professional, or personal purposes. OSN with a more professional and business-oriented background are typically used as a facility geared towards career management or business contacts; such networks typically provide SNS with a more serious image. In contrast, OSN with a more private and leisure-oriented background are typically used for sharing and exchanging more personal information, like, e.g., contact data, photographs, and videos; OSN provided by such networks have usually a more youthful interface. The core OSN application is the creation and maintenance of *contact lists*.

[1] According to reports, facebook.com recently surpassed the previously most popular website google.com by both page visits and served bandwidth:
http://www.hitwise.com/us/datacenter/main/dashboard-10133.html
http://www.mercurynews.com/business/ci_14698296?nclick_check=1.

This work has partially been funded by ETRI, DFG FOR 733 ("QuaP2P"), and the EU SOCIALNETS project, grant no 217141.

Through informing users automatically on profile changes of their contacts, the SNS thus helps users to stay up to date with news of their contacts and very often the popularity of users is measured in the size of their contact lists.

These properties of the SNS have led to the definition of boyd and Ellison [6], according to which *Social Network Sites* or *Online Social Network Services* are:

> *" ... 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".*

This definition, however, leaves aside some additional services that become apparent when observing the use of SNS. In particular, the communication of members through direct, sometimes instant message exchange, the annotation of profiles (e.g., via comments and recommendations), or the creation of links pointing to other profiles (picture tagging). The publication and browsing of images has grown to become a core function of these services [13]. Additionally, SNS typically provide support for a variety of third-party applications featuring advanced interactions between members ranging from simple "poking" of another member or the support for interest groups for a common topic to "likeness" testing with other members and the exchange of virtual gifts.

Maintenance and access to the OSN and their services are offered by commercial ***Social Network Providers*** (SNP), like Facebook Inc.[2], LinkedIn Corp.[3], Google Inc.[4], XING AG[5], and the likes. In general, a large amount of PII provided by the users is stored at the databases being under control of these providers, especially in the case of OSN targeting non-professional purposes. This data is either visible to the public, or, if the user is aware of privacy issues and able to use the settings of the respective SNS, to a somewhat selected group of other users. As profiles are attributed to presumably known persons from the real world, they are implicitly valued with the same trust as the assumed owner of the profile. Furthermore, any actions and interactions coupled to a profile are again attributed to the assumed owner of this profile, as well.
A SNP can, together with its SNS, also offer an application programming interface (API), allowing interested users to program a ***Social Network Application*** (SNA), thus extending and enhancing the functional range of the service.

Unfortunately, the popularity and broad acceptance of social networking services as platforms for interaction and social activities attracts not only faithful users, who are trying to add value to the community, but parties with rather

[2] www.facebook.com
[3] www.linkedin.com
[4] www.orkut.com
[5] www.xing.com

adverse interests, be they commercial or plain malicious, as well. Analyzing the OSN with respect to their security properties and the privacy of their users exposes some obvious threats. Different studies have shown that the participants clearly represent the weak link for security in OSN and that they are vulnerable to several types of social engineering attacks[6]. This partially is caused by a lack of awareness to the consequences of simple and presumably private actions, like accepting contact requests, tagging pictures, as well as acts of communication like commenting on profiles or leaving wall posts. However, the usability of privacy control mechanisms offered by the SNS and more importantly inherent trust assumptions on other users and their profiles, which are actually a desired social characteristic, certainly add to the problem.

The analysis of the privacy problems in current OSN demonstrates that even if all participants were aware and competent in the use of SNS, and even if a concise set of privacy measures were deployed, the OSN would still be exposed to potential privacy violations by either the omniscient service provider or an external attacker taking control of the OSN: the complete PII, directly or indirectly supplied by all participants, is collected and stored permanently at the databases of the providing company, which potentially becomes a *big brother* capable of exploiting this data in many ways that can violate the privacy of individual users or user groups.

The importance of this privacy exposure is underlined by the market capitalization of these providers, which ranges from 580 million US$ (acquisition of myspace through the news corp. in 2005) to 15 billion US$ (Facebook Inc, according to the investment of Microsoft in 2007)[1]. Even considering the commercial bodies that act as SNP to be trusted, hackers may be able to compromise their systems to gain access, unsatisfied employees may abuse their access to the data, or even imprudent publication of seemingly anonymized data may lead to the disclosure of PII, as it has happened in the past[7]. In consequence, we consider the protection of PII in OSN as an emerging topic, which is currently not addressed by the providers in the appropriate way.

1.1 Social Network Providers and Their Customers

Social network providers offer social networking services to the users and may further provide additional interfaces and services to other customers. These customers may come from different domains and pursue various goals.

In particular, ***sponsors*** belong to customers who advertise their services to the users through the OSN platform. Their advertisements may be of different

[6] Several of these attacks have been shown to be successful in the past. A short selection of examples can be found in [9, 5] as well as at `http://www.sophos.com/pressoffice/news/articles/2007/08/facebook.html` and `http://www.blackhat.com/html/bh-usa-08/bh-usa-08-archive.html`

[7] http://www.nytimes.com/2006/08/09/technology/09aol.html

kinds: plain commercial sponsors buy banner space or other marketing services from the SNP to advertise their products; SNS frequently contain "market pages" at which users can publish classified ads, job offers, and the likes, for which they may be billed. Also sponsors may create commercial interest groups or profiles inside the OSN.

Another type of OSN customers are ***third party service providers***, who extend the content and functionality of SNS with their own applications. These applications such as quizzes and games are typically executed on the servers under control of these third parties connected to the SNS via appropriate APIs. Often these applications have extensive access to the personal data of OSN users.

Finally, all sorts of ***data analysts*** may act as customers of SNP. These customers typically have data mining interests and may also get access to the personal information of users and their activities within the OSN. The analysis carried out by data analysts may serve different purposes, including scientific research (such as statistics, social behavior, or network-relevant aspects) and non-scientific data mining, typically for commercial purpose such as marketing.

Figure 1 illustrates the diversity of OSN customers and reflects their relationship to the SNS functionality and possible access to the personal information of the OSN users.

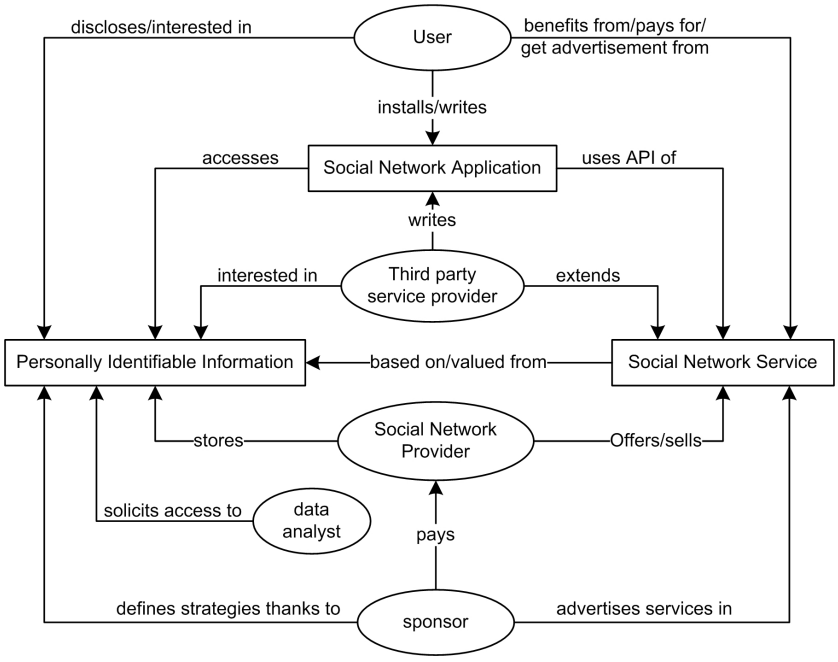

Fig. 1. OSN customers and their relationships to PII and SNS

1.2 Functional Overview of Online Social Networks

Even though each OSN is usually tailored to some specific use, the functional range of these platforms is essentially quite similar. Generally speaking, OSN functionality can be classified into three main types: The ***networking functions*** serve the actual purpose of OSN to foster social relationships amongst users within the virtual platform. In particular, they provide functionality for building and maintaining the social network graph. The ***data functions*** are responsible for the management of user-provided content and communications amongst the users. Their variety contributes to the enhancement of users' interaction and makes the platform more attractive. Finally, the ***access control functions*** aim to implement the user-defined privacy measures and to restrict unauthorized access to the user-provided data and information. In Figure 2 we illustrate the functionality provided by a typical OSN platform and provide more details thereafter.

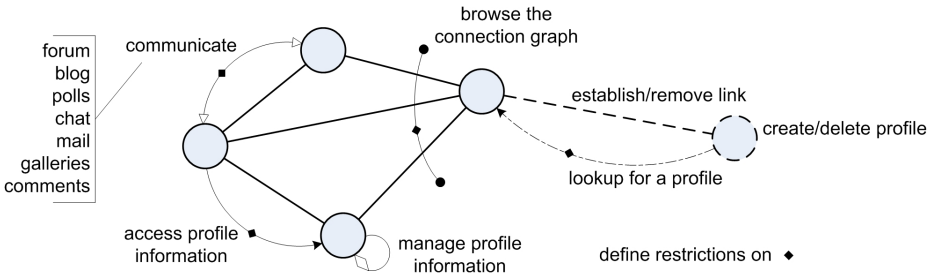

Fig. 2. Main functionality of a typical OSN platform

Networking functions. OSN users can typically build their profiles and establish relationships with each other. The set of networking functions includes all functions that update the vertices and the edges of the social network graph. In particular, the OSN user invokes the *profile creation* function upon his or her registration on the OSN platform. This function adds a new vertex representing that user to the social network graph. Thereafter, with *profile lookup* the user can find other users, who are also represented via vertices. Through the call to the *relationship link establishment* function the user can set up a new relationship with some other user. This function typically sends notification to that user, who in turn can accept or ignore the request. If the user accepts the request then users are added to the contact lists of each other and a new edge representing their relationship is added to the social network graph. The OSN users can also encounter profiles for possible relationships thanks to the *contact list browsing* function, which is realized through the traversal along the edges of the graph. Additional networking functions can be used to remove vertices and edges from the graph, for example upon the deletion of the user's profile.

Data functions. OSN users can typically advertise themselves via their own profiles and communicate with each other using various applications like blogs, forums, polls, chats, e-mails, and online galleries. Here we point out the *profile update* function, which allows the OSN users to maintain details on their own profiles and provide fresh information to other users, who may call the *profile retrieval* function, and hence visit the profile. Communication amongst users via blogs and forums is typically implemented through the *post* function, which inserts a block of information as an element into the main thread (sometimes called the "wall"). This block of information is not limited to plain text and can also contain videos, pictures, or hyperlinks. An OSN user willing to setup multimedia galleries typically calls the *upload* function, which transfers digital data from user's device to the OSN database. In case of content depicting other users, the *tag* function can create a link pointing to their profile. OSN users can typically evaluate content published by other users through the *like* or *dislike* functions. These functions can also be considered as a feedback to the publisher given by other users. In consequence, the user may either be encouraged, or discouraged to provide similar uploads and posts. Using the *comment* function OSN users can articulate their point of view in a more explicit way. OSN users can also exchange personal messages. Here, in particular, the *write to* function simulates the asynchronous offline communication (e.g., e-mail), whereas the *chat to* function allows for the synchronous real-time communication. An OSN user can send messages to individuals and also to subgroups of users from his or her contact list. The latter subgroup can be defined via the *regroup* function. Additionally, users may *create* interest groups, *advertise* own interest groups to other users, and *join* interest groups created by other users. The user who creates an interest group obtains administrator rights for this group by default; however, these rights can be changed thereafter, and distributed to other group members.

Access control functions. OSN users are usually allowed to define their own privacy settings through the some control functions. In particular, an OSN user may have control over the

- visibility of the online presence within the OSN
- visibility of contacts from the user's contact lists
- visibility and access to his or her own profile information
- access to his or her own uploaded content and posted communications

All these functions usually take as an input the information to be protected and the list of profiles having the rights to access it. The eligible profiles can be clustered into generic groups such as "friends", "friends of friends", "everybody", or user-defined groups, such as "family", "colleagues" and the like.

For example, the profile lookup function takes as an input a target's profile identifier, such as the name of the profile owner, and returns a list of possible candidates. An OSN user can apply output restrictions on this function to partially hide the own presence in the OSN. However, the protected profile would

remain reachable due to the profile browsing functionality of the OSN. Nevertheless, sensitive relationships can be hidden from unauthorized users by imposing restrictions on the output of the contact list browsing function. Thus, combined with the restrictions on profile lookup, this constraint can completely hide some profile in the OSN, since this profile will become unreachable from other users outside of the profile's contact list. Note that new contacts could be still added to the profile owner's contact list on the initiative of the latter. Another example is the control on the output of the profile retrieval function, which allow the profile owner to control the disclosure of the profile to other users. This allows some OSN user to hide parts of the private profile information from selected partners. Finally, the data related to online or offline indicators, one-to-one or one-to-many communications, such as posts, walls, comments, positive or negative marks, tags and the like can be protected by the means of restrictions on the huge set of the communication functions.

1.3 Modelling Data contained in Online Social Networks

The core information stored in OSN, the self generated and maintained data of the users and their profiles, can be classified into the following five types (cmp. Fig. 3):

1. Personal contact details, describing the user's identity
2. Connectivity, representing the connections in the social network graph
3. Interests of the user
4. Information on the curriculum vitae of the user
5. Communication, including all interactions with other OSN users of the SNS

These types encompass the amount of personally identifiable information, which is provided directly by the OSN user. Additional information about the OSN user is often generated and made accessible within the OSN by other users.

Personal contact details describe *'who the user is'*, providing not only some basic information such as user's name, picture, gender, birthday and birthplace, and marital status, but also some additional meta information with regard to the membership in the OSN, as well as the contact information aside of the OSN platform, such as (e)mail addresses, phone numbers, instant messaging identifiers, and personal web sites. Furthermore, it describes the personal profile of the user and may report about sexual, personal, political or religious interests and preferences. Users frequently can include a quick summary about themselves, describing their professional expertise, views and opinions, skills they "have to offer", as well as a short text on what they are looking for.

Connectivity describes *'whom the user knows'*, providing the user's contact list, possibly with annotated information about the type of the relationship (cf. family, colleagues, best friend, sports partner). Especially OSN platforms that

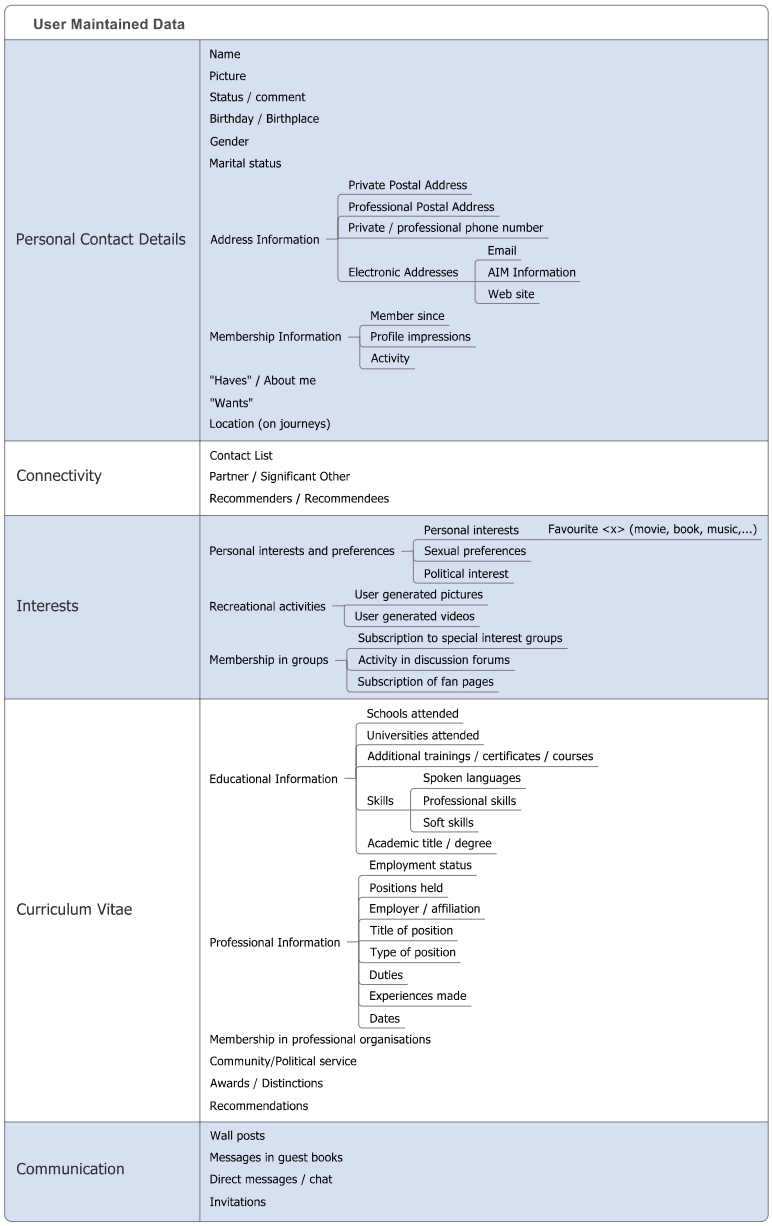

Fig. 3. Types of data commonly stored in OSN profiles.

with more private and leisure-oriented focus frequently ask the user to provide information on the relationship status, and in consequence the name and profile of their significant other. Users may further ask for recommendations by others.

These recommendations may contain very detailed information about the user, and shed light on the relationship between the both.

Interests describe *'what the user likes and is interested in'*. These may contain user's personal interests, hobbies, and preferences: In particular, information about favorite movies or music style, their sexual, religious, and political views, recreational activities of the user (such as personal pictures and videos showing situations from their personal lives), and their subscription to fan-pages as well as membership in special interest groups inside the OSN (which usually can be resolved to reading their posts containing their opinions on different topics).

Information on the curriculum vitae of the user describes the *professional career* and *educational background*, including attended schools, colleges, and universities, advanced studies, academic titles and professional certificates, as well as professional and soft skills. This information may be very detailed and include the description of job positions the users currently hold or have previously had, usually including information on the duration and type of the position (e.g. full-time, part-time, freelance, self-employed), the duties and responsibilities fulfilled in the job, and experiences being collected.

In addition to this description of the career progression, some OSN platforms ask the users to provide information on their membership in professional organizations (past and present), their community and political services (memberships and positions in clubs, associations, political parties, and professional societies), awards and distinctions, as well as recommendations and references.

Communication describes *'which messages the user has exchanged and with whom'*. OSN platforms generally offer exchange of personal offline messages, asynchronous communication via posts on walls and guest books entries, which the profile owner may hide or disclose to other users, and synchronous communication such as chats. These are examples of direct communications initiated by the user. However, there are also some less direct communications provided by other functionalities of the OSN platforms, such as the utilization of SNS applications (e.g. "poking", "likeness tests", quizzes), as well as public or targeted invitations to organized events.

Indirect information disclosure about OSN users may occur through posted opinions and comments, or any type of annotations to profiles of other users. Even though the owners of the annotated profiles may be able to remove undesired annotations, they need to notice the annotations in the first place. Since many users do not explicitly search for annotations made by other users about their profiles, this indirectly disclosed information may remain publicly accessible over a longer period of time. Similarly, information about users may be disclosed

via third party statements about the user made in forums of the interest groups, or as annotations or comments at the profiles of other users.

Any form of user-generated digital content may also cause third party information disclosure. For example, some OSN networks try to prevent users from posting photographs showing people on their profiles if the owner of the profile is not pictured there[8]. However, this does not prevent users from posting photographs picturing them together with others. Additionally, many OSN platforms offer "tagging" of pictured users, whose profiles will usually be directly linked to that picture. These tags may contain further comments added by the user who uploads the picture.

1.4 A Model for Social Network Services

Social Network Services (SNS) are structured along the following three layers with different responsibilities (see also Figure 4):

- a ***Social Network*** (SN) level, building the digital representation of members and their relationships;
- a ***Application Service*** (AS) level, constituting the application infrastructure managed by the SNS provider;
- a ***Communication and Transport*** (CT) level representing the communication and transport services as provided by the network.

The SN layer provides each member with a set of functions corresponding to social interactions in the real life. These functions can be divided in two classes. The first class deals with *Communication Management* and includes the real-time communication, such as chat and phone calls, as well as the offline communication, such as wall posts, mails and tweets. The second class deals with *Relations Management*, including friendship requests, friends lookups, profile access and reputation administration. To implement these functions, the SN layer relies on the AS layer. This layer includes either the whole infrastructure managed by the SNS provider and the web, storage, and communication services to create the SN service. The AS layer can implement data storage and its retrieval, indexing of the content, management of access permissions to data, and node join or leave, in a centralized or distributed fashion. In any case, redundancy and delegation are common strategies to enhance availability: both for organizational reasons or if a server faces failures or other inabilities to provide the service, it may delegate requests to secondary or fallback servers. The AS layer in turn relies on the transport and (inter)networking protocols and infrastructures, such as the classical Internet or the GPRS connectivity. This network infrastructure is thus implemented by the lowest CT layer and can be managed by one or more network providers.

[8] `http://www.odnoklassniki.ru`

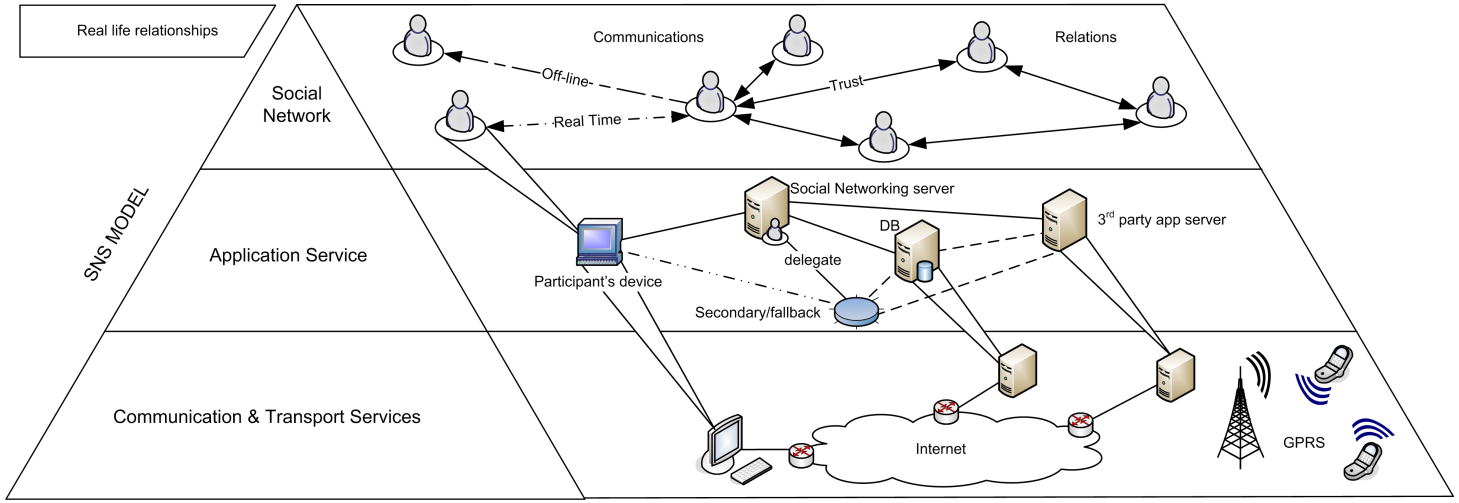

Fig. 4. Three architectural layers of social networking services.

2 Security Objectives: Privacy, Integrity, and Availability

Security objectives are requirements that have to be satisfied in order to protect the system from potential threats and attacks.

In this chapter we provide an overview of important security objectives for online social networks. First of all we notice that classical requirements (cf. [3]) of *confidentiality*, *integrity*, and *availability*, have a special touch when considered in the scope of OSNs. While integrity and availability have only subtle differences compared to other communication systems, in that they mostly address the content provided by the users, the requirement of confidentiality (usually associated with encryption) is no longer sufficient and should be extended to the more comprehensive security objective — *privacy*.

While potential breach of user privacy and integrity of user-provided contents may lead to economic damages for the users, cause embarrassing situations, and also tarnish their reputation (even in the real world), the missing availability of contents or services may also decrease the attractiveness of the actual OSN platform and harm its provider. It is extremely difficult to cope with all these goals simultaneously. Especially privacy of OSN users is challenging since the amount of personal information is huge and this information may be available not only from a particular OSN platform but also from the web.

In the following, we describe privacy, integrity and availability objectives for online social networks, while also mentioning potential threats with regard to not only the profile owner, but also other users and the system itself.

2.1 Privacy

Privacy is a relatively new concept, born and evolving together with the capability of new technologies to share information. Conceived as 'the right to be left alone'[15] during the period of newspapers and photographs growth, privacy now refers to the ability of an individual to control and selectively disclose information about him.
The importance of privacy is so relevant to have been reported in the Universal Declaration of Human Rights (art.12):

> *"No one shall be subjected to arbitrary interference with his privacy, family, home or correspondence, nor to attacks upon his honor and reputation. Everyone has the right to the protection of the law against such interference or attacks."*

In the internet age, where huge amount of sensitive data can be easily gathered, stored, replicated and correlated, the protection of privacy is even more seen as the main objective for the services provided by an OSN platform [8, 2].

Generally speaking, the protection of information which users publish at their profiles, presumably accessible by their contacts only, takes place with the *usage control*[12]: the control of the degree at which sensitive data is disclosed to selected other parties (i.e. access control) together with the control of its later

usage, even after the information has been accessed. Access to the contents of a user profile may only be granted by the user directly, and this access control has to be as fine-grained as the profile itself. For example, if the profile contains several information blocks then access to each block has to be managed separately.

In addition, the protection of communication calls for inference techniques aiming at deriving any type of information with regard to: (1) *anonymity*, meaning that users should access resources or services without disclosing their own identities; (2) *unobservability*, i.e. the requirement that no third party should gather any information about the communicating parties and the content of their communication; (3) *unlinkability*, which requires that obtaining two messages, no third party should be able to determine whether both messages were sent by the same sender, or to the same receiver; (4) *untraceability*, which demands that no third party can build a history of actions performed by arbitrary users within the system; in other words, it demands both anonymity and unlinkability.

In summary, the objective of privacy is to hide any information about any user at any time, even to the extent of hiding their participation and activities within the OSN in the first place. Moreover, privacy has to be met by default, i.e. all information on all users and their actions has to be hidden from any other party internal or external to the system, unless explicitly disclosed by the users themselves.

2.2 Integrity

The objective of integrity in online social networks is to prevent any unauthorized modification or tampering of user-generated content and profile information, as listed on Figure 3. This encompasses the protection of real identity of users within the OSN platforms. In this sense, the meaning of integrity in such networks is somewhat extended in comparison to the conventional detection of modification attempts on data. Moreover, problems with integrity of user profiles and their contents may have devastating impact on the objectives put forth with respect to the privacy of OSN users. Since the creation of profiles in traditional OSNs is easy, it is a matter of facts, that protection of real identities is insufficient in today's platforms. In particular, none of the current major OSN providers is able (and perhaps even not interested in) to ensure that a profile is associated to the corresponding individual from the real world.

As users inherently trust the OSN providers, the aforementioned vulnerabilities can be thwarted through the appropriate authentication procedures to assure the existence of real people behind registered OSN profiles. Identity checks do not necessarily have to be performed by a centralized service, however, all identification services have to be trusted by all participants.

2.3 Availability

The objective of availability for online social networks aims at assuring the operability of the social network services in the face of attacks and faults. The insufficient guarantees for availability may prevent users from accessing the service and make of the OSN platform less attractive. Especially, for OSNs with professional focus, e.g. OSNs that aid their users to foster business relations or find new job positions, it is mandatory to keep users' data continuously available. Therefore, we consider availability of user-generated data and profiles as a basic requirement that should be provided by the platforms, even though for leisure-oriented OSN platforms the availability of certain content may appear not of prime importance at first sight.

The main concern of availability are ***denial-of-service*** attacks. In the context of social network services they may aim at either seizuring a victim's profile (or selected parts of it) or disrupting the possibility to communicate with the user. Furthermore, also integrity threats like data pollution and cloning may impair the availability of network services by affecting the quality of the service perceived by the users.

Remark 1. Also distributed services, which are implemented in a decentralized way, possibly via peer-to-peer systems, or which follow other types of service delegation, may be vulnerable to a series of attacks against availability as well. These attacks include *black holes*, aiming at collecting and discarding a huge amount of messages; *selective forwarding*, where some traffic is forwarded to the destination, but the majority is discarded; and *misrouting*, which aims to increase the latency of the system or to collect statistics on the network behavior. In any case, attacks on distributed social networks are more effective in case of *collusion* amongst malicious users or in the presence of Sybil nodes controlled by the attacker, which is not the case for the centralized OSNs.

Finally, we notice that while privacy has to address broader spectrum of threats and deal with different types of attackers, including the OSN and application providers, as well as external parties, both integrity and availability primarily address the latter, since OSN users have an inherent interest that these objectives are met.

3 Attack Spectrum and Countermeasures

The diversity of available OSN platforms opens doors for a variety of attacks on privacy of the users, integrity of their profiles, and the availability of the user-provided contents. In this section we will highlight main attack types against OSN platforms and discuss their impact on the aimed security objectives. Table 1 will serve as a background for our discussion. It illustrates different types of attacks and shows their relevance for the mentioned security objectives of privacy, integrity, and availability. We will discuss not only the purpose and

	Security Objectives		
	Privacy	Integrity	Availability
Attacks			
Plain Impersonation	x	x	
Profile Cloning	x	x	
Profile Hijacking	x	x	
Profile Porting	x	x	
Id Theft	x	x	x
Profiling	x		
Secondary Data Collection	x		
Fake Requests	x		
Crawling and Harvesting	x		
Image Retrieval and Analysis	x		
Communication Tracking	x		
Fake Profiles and Sybil Attacks		x	
Group Metamorphosis		x	
Ballot Stuffing and Defamation		x	
Censorship		x	x
Collusion Attacks	x	x	x

Table 1. Attacks vs. Security Objectives in Online Social Networks

impact of each attack but also explain the techniques needed to mount it, while referring to some real-world examples, where possible. We note, however, that technical realization behind an attack may strongly depend on the functionality and in particular on the use of different protection mechanisms within the OSN platform. Therefore, not every attack technique will have the same impact when used against different OSN platforms. Moreover, since OSN providers typically have full control over the network resources, no meaningful protection appears possible if the attacks are mounted by the provider itself.

3.1 Plain Impersonation

With *plain impersonation* attack the adversary aims to create fake profiles for real-world users as depicted on Figure 5. In this sense a real-world user will be impersonated within the OSN platform. The success of this attack strongly depends on the authentication mechanisms deployed in the registration process. Since many OSNs tend to authenticate email addresses by requesting confirma-

tions for the registration emails, this attack can be easily performed if an email address is created in advance. The consequence of plain impersonation is that the adversary can participate in the OSN applications on behalf of the impersonated user with all damaging consequences for the user. A currently very prominent secondary effect of all kinds of impersonation (Sections 3.1 – 3.5) is the misuse of the trust that users inherently have in messages from their accepted contacts, and especially the "419" scam[9]: impersonating attackers engage in a dialog with contacts of the impersonated individual, and, by producing a credible story, ("My wallet was stolen in London and now I can't pay my flight home") successfully defraud the victim. This attack can be thwarted only through the deployment of stronger authentication techniques. In particular, it is desirable to require some form of real-world identification from the user prior to switching on her account.

3.2 Profile Cloning

By *profile cloning* we understand a special type of impersonation attack that occurs within the same OSN platform [5], as depicted on Figure 5. The goal of the adversary here is to create a profile for some user that is already in possession of some valid profile in the same network. From the technical point of view this attack can be realized through the registration of the new profile using the same (or similar) content as the existing one. This is feasible in most OSN platforms since each profile is associated with some unique administrative id and an email address used during the registration. Furthermore, many users hide their email address so that OSN users would not be able to distinguish between the original profiles and their clones registered with other email addresses. As a consequence the adversary can create confusion through impersonation of other registered users and possibly gain access to the private information communicated to that users. Moreover, with tools like iCloner [5] profile cloning can be automated. Such tools are able to collect public data of OSNs members, match them, create cloned profiles and then send friendship requests on their behalf. A possible solution for OSN providers to prevent profile cloning is to deploy mechanisms that are able to detect similarities between different profiles, in particular with regard to the personal information that is visible to the OSN users. Since cloned profiles typically have later registration date than the original ones, it should be feasible for the OSN provider to distinguish them and remove from the network.

3.3 Profile Hijacking

The goal of the adversary mounting a *profile hijacking* attack is to obtain control over some existing profile within an OSN platform. Many OSN platforms protect user access to their own profiles via passwords. Hence, from the technical point of view profile hijacking is successful if the adversary can obtain passwords of other users. This can be done by many means. First, it is a well-known fact that the majority of users choose weak passwords that can be recovered via

[9] `http://www.419scam.org/`

an automated dictionary attack [7]. However, OSN providers typically deploy protection against such attacks by restricting the number of login attempts or by using techniques that require human interaction such as CAPTCHAs [14]. Nevertheless, there exist effective tools, e.g. as the one included in iCloner [5], that are able to analyze and bypass CAPTCHAs. Alternatively, the adversary may try to obtain passwords via social-engineering attacks such as phishing [11], or obtaining passwords for other online services, relying on the fact that most people use the same passwords across the majority of their accounts at different sites. The OSN functionality can be misused to distribute messages aiming to lure users to fake login websites[10]. Finally, we shouldn't forget that OSN providers themselves have full control over the registered profiles. Therefore, if some profile appears attractive for the OSN provider to be hijacked the password access to the profile can be changed accordingly.

3.4 Profile Porting

By *profile porting* we understand another type of impersonation where some profile that exists within one OSN platform is cloned into another OSN platform [9, 5], as depicted on Figure 5. From the technical point of view this attack can be realized via registration of a profile using some new email address. Profile porting is appealing since not every user has her own profile on every available OSN platform. On the other hand, there might be some users that participate in both OSN platforms and thus will not be able to distinguish amongst ported profiles. The significance of profile porting (e.g. in comparison to profile cloning) is that users may be completely unaware that their profiles have been ported. The impact of profile porting is that the adversary can impersonate users in different OSN platforms. Thwarting profile porting is not that easy. In particular, profile similarity detection tools can still be used but only if they can work across multiple OSN platforms. Since every OSN platform is administrated by a different provider, the deployment of such tools would require cooperation amongst the providers. This is difficult to achieve, since OSN providers are cautious about granting any form of access to their profile database to competitors.

3.5 ID Theft

Under *ID theft* we consider the impersonation of OSN users in the real-world [5], as depicted on Figure 5. An adversary mounting the ID theft attack should be able to convince anyone about the ownership of some particular OSN profile. In this way the adversary can possibly misuse the reputation or expertise of the real profile owner for own benefit, while leaving the owner unaware of the attack. One way for a successful ID theft attack is to take control over the target profile. This requires the same effort as for the profile hijacking attack. However, this effort seems necessary only if the adversary has to actively use the

[10] `http://fraudwar.blogspot.com/2009/05/facebook-hack-reveals-trend-in.html`

profile for the ID theft attack, e.g. communicate via the OSN platform. Often it would simply suffice to claim the ownership of a profile and perform the actual communication via other channels. In this case thwarting ID theft attacks by technical means seems impossible. The only solution is to rely on other means of real-world identification such as national identity cards, driver's licenses, etc.

3.6 Profiling

In addition to the maintenance of own profiles modern OSNs provide users with various applications to express themselves via forums, guest books, discussions, polls, multimedia data, etc. These activities are observable by other users within the OSN platform. By *profiling* we understand an attack against any target OSN user aiming to collect information about OSN activities or further attributes of that user, e.g [4], see also Figure 6. This attack can be typically performed by OSN users, possibly in an automated way, since the collectable information is usually publicly accessible by all OSN users. The risk of profiling attacks performed by OSN users can be diminished via fine-grained access control and anonymizing techniques. For example, users should be able to allow access to the personal parts of their profile on the individual basis and not only based on roles (e.g. friends) as realized in many current OSN platforms. However, the recent studies, e.g. [10], show that even if the personal information is hidden, it can still be inferred from public information and social activities of the user. An alternative solution could be to let users decide whether their activities (e.g. discussion comments) should be kept unlinkable to their profiles. Although these measures may help to reduce the risk of profiling performed by other OSN users, thwarting profiling performed by OSN providers[11] appears to be much more difficult.

3.7 Secondary Data Collection

By *secondary data collection* we understand an attack that aims to collect information about the owner of some OSN profile via secondary sources apart of the OSN platform, as depicted on Figure 6. A typical example of secondary data collection is to use some Internet search engine to find information that can be linked to the profile owner. More effective is to use some Internet service[12] that aggregates all information it can find about some particular person. Through such an attack the adversary may obtain much more information about some user than available in the profile and misuse it against the user both in the virtual environment of the OSN platform and in the real life. Another example are recent de-anonymization attacks [16] that misused the group memberships of social network users for their unique identification. Furthermore, the existence of OSNs with public and private profiles simplifies the secondary data collection as many users tend to have accounts on different platforms [18]. There is no

[11] `http://www.pcworld.com/article/191716/myspace_user_data_for_sale.html`
[12] `http://www.123people.com/`

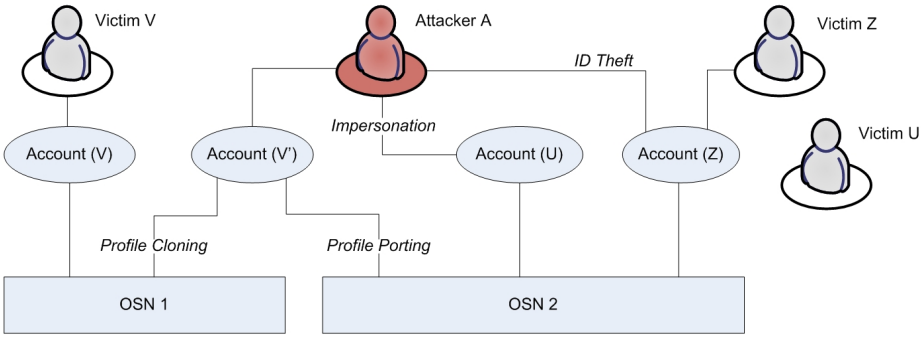

Fig. 5. Impersonation attacks: victim U doesn't have any OSN account, victim V has an account on OSN1 and victim Z on OSN2. The attacker V generates U's account on OSN1, a copy of V's account on OSN1 and OSN2, and logs on OSN2 with the credentials of Z.

meaningful protection against secondary data collection attacks since the data is typically aggregated from different locations. Therefore, it appears in responsibility of the user to limit information kept in the profile in order to avoid its linkability with secondary sources.

3.8 Fake Requests

One of the main objectives of OSN platforms is to establish social contacts. This proceeds via connection requests that can be either accepted or rejected by the users. An adversary with own OSN profile that sends *fake requests* to other users aims less on the social contact with these users but is more interested to expand its own network. The dissemination of fake requests can be automated. Since many OSN users tend to accept fake requests[13], the adversary can simplify access to their profiles and activities and possibly obtain additional information, whose visibility is subject to the available direct or nth-grade connections. These connections can then be misused for the automated collection and aggregation of information. The actual dissemination of fake requests cannot be prevented since establishment of new connections is an important goal of OSN applications. Therefore, it is desirable that users behave more responsibly upon accepting new connection requests. Unfortunately, current studies, e.g. [5] show that users tend to accept fake requests.

3.9 Crawling and Harvesting

The goal of *crawling* is to collect and aggregate publicly available information across multiple OSN profiles and applications in an automated way [5, 4]; see also

[13] `http://www.columbiamissourian.com/stories/2005/09/01/a-new-kind-of-fame/`

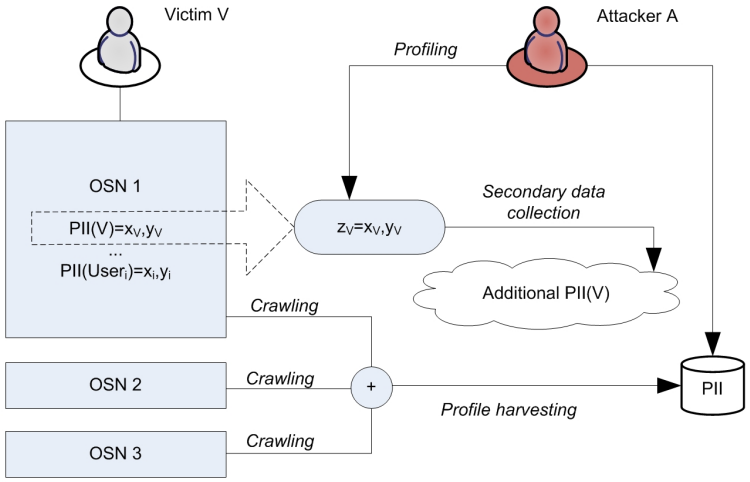

Fig. 6. Main PII related threats in current OSNs.

Figure 6. Unlike profiling this attack does not target any particular user and unlike secondary data collection it is executed within the OSN environment. The expansion of own network connections by the adversary using fake requests can be seen as a preliminary step for crawling. The adversary is simply interested in collecting as much public information within the OSN platform as possible. This information can then be misused for different purposes, for example for selling data to marketing agencies, etc. Also it would allow for the offline analysis of social relationships and user activities, thus paving the way for targeted attacks on OSN users. Although some OSN platforms try to protect from crawling through the deployment of CAPTCHAs, the latter can be passed over with the appropriate solving tools [5]. Another attack by which the adversary simultaneously crawls across different OSN platforms is called *harvesting*. Typically harvesting results in larger datasets with larger amount on private information about the OSN users.

3.10 Image Retrieval and Analysis

Upload of images or other digital content and its discussion stimulates social interactions of OSN users. However, free accessibility to images and videos bear potential risks to the privacy of users. By *image retrieval and analysis* we understand an automated attack aiming to collect multimedia data (incl. images, videos, etc.) available with the OSN platform. This attack is typically followed by the subsequent analysis via automated pattern recognition tools (see e.g. [17] for a survey on face recognition) to find links to the OSN profiles of displayed users. Information distilled in this way can reveal more private information about users than they are willing to give. In particular, it may reveal information about friends or colleagues that are not necessarily part of the user's

social network, or information about visited locations (location-tracking) shown on the photographs. The analysis of digital content can be further strengthened by considering secondary sources such as search over the Internet. Digital content retrieval attacks can be possibly thwarted through a more restrictive access control policies for the digital content.

3.11 Communication Tracking

OSN users communicate with each other using diverse OSN applications. By *communication tracking* we understand a profiling attack aiming to reveal information about communications of the same user. In this way the attacker may collect more information about the user than available in the profile. This attack can be mounted in an automated way by searching for comments left by the target user in various OSN applications.

3.12 Fake Profiles and Sybil Attacks

In many OSN platforms users can easily create several profiles under possibly different identities and contents. Since many OSN platforms lack of proper authentication such creation of *fake profiles* becomes easy [5]. On the technical side, the user has only to create a new email for the registration of a fake account. Fake profiles pave the way for *Sybil attacks* that may serve different purposes[14,15]. For example, owners of fake profiles can establish new connections without disclosing their real identities. In this way they may obtain more information about some person than by using some real account. Sybil account may also be created on behalf of the whole groups[16]. Furthermore, Sybil accounts can be misused against the functionality of the OSN platforms. This includes distribution of spam messages[17] or other illicit content such as malware[18] and phishing links[19,20], illegal advertisement, bias of deployed reputation systems, etc. Creation of fake profiles can be seen as a special form of impersonation attacks. One solution for OSN providers to recognize fake profiles is to use IP traceback. Indeed, if logins to several profiles come from the same IP address then it is likely that some of these profiles are fake. However, an attacker may try to avoid IP traceback by using different proxies. Therefore, stronger identification and authentication mechanisms for admission of new users would offer a better protection.

[14] http://www.nature.com/news/2009/090423/full/news.2009.398.html
[15] http://www.sophos.com/pressoffice/news/articles/2009/12/facebook.html
[16] http://gadgetwise.blogs.nytimes.com/2010/03/18/fake-facebook-fan-pages/
[17] http://www.pcworld.com/businesscenter/article/191847/facebook_users_targeted_in_massive_spam_run.html
[18] http://content.usatoday.com/communities/technologylive/post/2009/12/koobface-compels-facebook-victims-to-help-spread-worm-/1
[19] http://scitech.blogs.cnn.com/2010/03/19/facebook-responds-to-massive-phishing-scheme/
[20] http://www.pcworld.com/businesscenter/article/174607/twitter_warns_of_new_phishing_attack.html

3.13 Group Metamorphosis

A popular application provided by OSN platforms is the establishment of shared interest groups. These groups are usually administrated by OSN users and provide a platform for more focused discussions, specialized contact establishment, and dissemination of information, which may be interesting for a targeted audience. By *group metamorphosis* we understand an attack where group administrators change the group subject to persuade own interests, e.g. political[21]. Other OSN users who joined the group earlier may remain unaware of this change, which in turn may have negative impact on their reputation. A possible solution for OSN providers to thwart group metamorphosis attacks is to restrict control of administrators over the interest groups, in particular to prevent them from modifying any information that may have impact on the group as a whole.

3.14 Ballot Stuffing and Defamation

OSN platforms serve primarily the contact establishment and interaction amongst users. Hence, attacks biasing public perception and recognition of a target OSN user by others are undesirable. By *ballot stuffing* we understand an attack by which the attacker wishes to increase public interest to some target OSN user. This attack may increase the amount of personal messages or connection requests received by the target user resulting in a DoS attack on the physical resources of the OSN user. The attack may place the victim into the focus of public, possibly embarrassing discussions. On the other hand, ballot stuffing may increase popularity of the profile belonging to the attacker. This can be achieved through recommendations submitted by the attacker using fake profiles. In contrast, *defamation attacks* aim at decreasing public interest of a target user, in particular by tarnishing the reputation of the latter[22]. In particular, defamation may lead to blacklisting of the user in contact lists of other users and keep the user away from participation in communication applications such as shared interest groups and discussion forums. It may further have negative impact on the user's life in the real world[23]. Another form of defamation is the anti-advertising against companies[24] aiming to damage the reputation of the latter on the market.

Both ballot stuffing and defamation attacks have to be performed at a large scale in order to have a significant impact. An attacker may create fake profiles and use automated tools to disseminate information needed to increase or

[21] One incident has been reported for facebook, where a multitude of groups have been fostered under general topics and concertedly renamed to support Silvio Berlusconi, in 2009 `http://www.repubblica.it/2009/12/sezioni/politica/giustizia-21/gruppi-facebook/gruppi-facebook.html`

[22] `http://timesofindia.indiatimes.com/sports/off-the-field/Rachel-Uchitel-threatenslawsuit-over-Facebook-defamation/articleshow/5708237.cms`

[23] `http://mybroadband.co.za/news/Internet/6580.html`

[24] `http://blogs.bnet.com/businesstips/?p=6786`

decrease interest to a specific OSN user. Another technique is to use the poll application provided by many OSN platforms and let users vote on information related to the victim.

3.15 Censorship

OSN providers typically have control over the whole data available within the network. As such they can deliberately manipulate the user-provided information and contents. In some cases this ability is necessary to prevent dissemination of illicit content. On the other hand, *censorship* when applied without substantial reasons may have negative impact on the OSN users. For example, in OSN platforms focusing on business contacts users often advertise their expertise. In this scenario censorship may be misused to favor some users over their competitors. Censorship may have many facets. It can be performed by active modification of user-provided contents, which might remain unnoticed by the user. Higher impact can be achieved through the target manipulation of search engines within the network. Since censorship can be performed by the OSN provider[25] without involving any other parties, there is little one can do to prevent this threat. Censorship may be applied not only by OSN providers but also by administrators of shared interest groups. They can deliberately modify or drop messages of group members. Although restricting group administrators from modification of other user contents appears to be an effective protection measure, it is unlikely to be used in practice, since this ability contradicts to the responsibility of group administrators for the content disseminated within the group.

3.16 Collusion Attacks

The "impact of a crowd" can be exhibited in OSNs through a *collusion* of users. In this attack several users join their malicious activities in order to damage other OSN users or mount attacks against applications of the OSN platform. In particular, colluding users may start defamation or ballot stuffing campaigns, increase each over reputations, bias the outcome of public polls or influence public discussions. Since colluding users have valid OSN profiles these attacks do not require creation of fake profiles. Furthermore, these attacks are more difficult to recognize than similar attacks mounted via fake profiles. The reason is that IP traceback would not help even if colluding users do not deploy any additional proxies.

4 Summary and Conclusion

This chapter deals with security in Online Social Networks (OSN). It introduces Online Social Networks as the digital representations of relationships, which their users entertain in the physical world. Social network providers (SNP), commonly

[25] `http://www.civic.moveon.org/pdf/myspace/`

commercial entities that offer Social Networking Services (SNS), the access to the OSN, and their users, are identified as the main actors in online social networking. Sponsors, application providers, and data analysts are third parties in this context and they represent further actors.

Analysing the typical users of OSN it becomes apparent that the seeming ease of use attracts especially individuals with limited knowledge about computers, the Internet, and computer security. The users provide and maintain a wealth of data to the OSN at the same time. Privacy concerns remain unaddressed, and the uploaded data largely consists of personally identifiable information (PII), and even private messages between the users.

In order to allow for the analysis of threats, this chapter attempts to formally divide typical SNS into a layered model of the Social Networking Layer, including the digital representation of their users' relationships, the Application Service Layer, comprising of the service infrastructure offering the social networking services, and the Communication and Transport Layer, which represents the underlying computer networks.

Comparing security objectives for online social networks to common security goals in computer science, which generally identified to be confidentiality, integrity, and availability, some similarities, and, more importantly, some differences become apparent.

Especially establishing confidential channels between senders and receivers is not sufficient for social networking services. The confidentiality objective hence has to be extended to provide *privacy* of the users. Considering the wealth of PII stored in OSN, and the extension through third party annotations, the accessibility of data more strictly has to be restricted in order to protect the identity and privacy of users. Threats to their users' privacy are abundant in OSN. The access to the data of users through the provided interfaces eases the automated gathering of data in order to profile single users, or to mine their data in multiple sources, thus allowing for secondary data collection. Even the harvesting of a large number of profiles becomes easily possible, thus collecting PII of not only particular, but even of large numbers of users.
Different types of identity threats are direct consequences, with impersonation, profile cloning, and even the porting of profiles being feasible and quite simple to accomplish.

The integrity, being threatened by creation of faked profiles, or even defamation and ballot stuffing, as well as availability, mainly threatened by denial-of-service and censorship, generally are quite similar to their traditional definitions in computer science.

Considering these objectives, a plethora of attacks on SNS security are conceivable, and many of these not only have been shown in scientific work, but reported to have been conducted in the wild, too. The attacks are classified into nine distinct groups and described in detail. Academic, and, where reported, real world examples are given and explained to further describe them and illustrate their impact.

Online social networks currently are among the best accepted and most highly utilized networked applications on the Internet. Their immense user base, which by 2010 has exceeded 500 million distinct users, permits predicting that they will stay being one of the killer apps during the coming years. However, both their users and their providers have not learned to properly master their properties, and the vast number and severity of threats, as well as the plethora of reported attacks, underlines the importance of introducing and enforcing security measures, which are better than the rudimentary approaches that are implemented today.

References

1. Modelling The Real Market Value Of Social Networks. Available at: `http://www.techcrunch.com/2008/06/23/modeling-the-real-market-value-of-social-networks/`, 2008.
2. danah m. boyd . Facebook's privacy trainwreck. *Convergence: The International Journal of Research into New Media Technologies*, 14(1):13 – 20, 2008.
3. A. Avizienis, J.-C. Laprie, B. Randell, and C. Landwehr. Basic concepts and taxonomy of dependable and secure computing. *IEEE Trans. Dependable Secur. Comput.*, 1(1):11–33, 2004.
4. M. Balduzzi, C. Platzer, T. Holz, E. Kirda, D. Balzarotti, and C. Kruegel. Abusing Social Networks for Automated User Profiling. Research Report RR-10-233, EURECOM, 2010. `http://www.iseclab.org/papers/socialabuse-TR.pdf`.
5. L. Bilge, T. Strufe, D. Balzarotti, and E. Kirda. All Your Contacts Are Belong to Us: Automated Identity Theft Attacks on Social Networks. In *18th Intl. World Wide Web Conference*, 2009.
6. d. m. boyd and N. B. Ellison. Social network sites: Definition, history, and scholarship. *Journal of Computer-Mediated Communication*, 13(1), 2007.
7. D. Florencio and C. Herley. A Large-Scale Study of Web Password Habits. In *16th International Conference on World Wide Web (WWW 2007)*, pages 657–666. ACM, 2007.
8. R. Gross and A. Acquisti. Information revelation and privacy in online social networks. In *ACM Workshop on Privacy in the Electronic Society*, pages 71 – 80, 2005.
9. T. N. Jagatic, N. A. Johnson, M. Jakobsson, and F. Menczer. Social phishing. *Communications of the ACM*, pages 94–100, 2007.
10. A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel. You Are Who You Know: Inferring User Profiles in Online Social Networks. In *ACM International Conference on Web Search and Data Mining (WSDM 2010)*, pages 251–260. ACM, 2010.
11. T. J. Nathaniel, N. Johnson, and M. Jakobsson. Social phishing. *the Communications of the ACM. Retrieved March*, 7, 2006.
12. J. Park and R. Sandhu. Towards usage control models: beyond traditional access control. In *SACMAT '02: Proceedings of the seventh ACM symposium on Access control models and technologies*, pages 57–64, New York, NY, USA, 2002. ACM.
13. F. Schneider, A. Feldmann, B. Krishnamurthy, and W. Willinger. Understanding Online Social Network Usage from a Network Perspective. In *ACM SIGCOMM conference on Internet measurement*, 2009.

14. L. von Ahn, M. Blum, N. J. Hopper, and J. Langford. CAPTCHA: Using Hard AI Problems for Security. In *EUROCRYPT 2003*, volume 2656 of *LNCS*, pages 294–311. Springer, 2003.
15. S. D. Warren and L. D. Brandeis. The right to privacy. *Harward Law Review*, 4(5):193–220, December 1890.
16. G. Wondracek, T. Holz, E. Kirda, and C. Kruegel. A Practical Attack to De-Anonymize Social Network Users. In *IEEE Symposium on Security and Privacy*. IEEE CS, 2010. `http://www.iseclab.org/papers/sonda.pdf`.
17. W. Zhao, R. Chellappa, P. J. Phillips, and A. Rosenfeld. Face Recognition: A Literature Survey. *ACM Computing Surveys*, 35(4):399–458, 2003.
18. E. Zheleva and L. Getoor. To Join or Not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *WWW 2009*, pages 531–540. ACM, 2009.