enisa
European Network and Information Security Agency

ENISA Position Paper No.1

Security Issues and Recommendations for Online Social Networks

Editor: Giles Hogben, ENISA

October 2007

Security Issues and Recommendations for Online Social Networks

ENISA position papers represent expert opinion on topics ENISA considers to be important emerging risks. They are produced by a group selected for their expertise in the area. The content was collected via wiki, mailing list and telephone conferences, edited by ENISA and the final version has been reviewed by the people listed below.

This paper aims to provide a useful introduction to security issues in the area of Social Networking, highlight the most important threats and make recommendations for action and best practices to reduce the security risks to users. Examples are given from a number of providers throughout the paper. These should be taken as examples only and there is no intention to single out a specific provider for criticism or praise. The examples provided are not necessarily those most representative or important, nor is the aim of this paper to conduct any kind of market survey, as there might be other providers which are not mentioned here and nonetheless are equally or more representative of the market.

Audience

This paper is aimed at corporate and political decision-makers as well as Social Network application-providers. It also seeks to raise awareness among political and corporate decision-makers of the legal and social implications of new developments in Social Networking technologies. In particular, the findings should have important implications for education and data protection policy.

Contributors

Alessandro Acquisti, Carnegie Mellon University
Elisabetta Carrara, ENISA
Fred Stutzman, UNC
Jon Callas, PGP Corp
Klaus Schimmer, SAP
Maz Nadjm, Rareface
Mathieu Gorge, Vigitrust
Nicole Ellison, MSU
Paul King, Cisco Systems
Ralph Gross, Carnegie Mellon University
Scott Golder, Hewlett-Packard

Group members participate as individuals. This paper should therefore not be taken as representing the views of any company or other organisation, and does not in any way bind group members when dealing with the issues it covers in other contexts.

Executive Summary

Online Social Networks or Social Networking Sites (SNSs) are one of the most remarkable technological phenomena of the 21st century, with several SNSs now among the most visited websites globally. SNSs may be seen as informal but all-embracing identity management tools, defining access to user-created content via social relationships.

Since the commercial success of an SNS depends heavily on the number of users it attracts, there is pressure on SNS providers to encourage design and behaviour which increase the number of users and their connections. Sociologically, the natural human desire to connect with others, combined with the multiplying effects of Social Network (SN) technology, can make users less discriminating in accepting 'friend requests'. Users are often not aware of the size or nature of the audience accessing their profile data and the sense of intimacy created by being among digital 'friends' often leads to disclosures which are not appropriate to a public forum. Such commercial and social pressures have led to a number of privacy and security risks for SN members.

This paper emphasises the commercial and social benefits of a safe and well-informed use of SNSs. It also outlines the most important threats to users and providers of SNSs and offers policy and technical recommendations to address them.

Threats

- **Threat SN.1 Digital dossier aggregation:** profiles on online SNSs can be downloaded and stored by third parties, creating a digital dossier of personal data.

- **Threat SN.2 Secondary data collection:** as well as data knowingly disclosed in a profile, SN members disclose personal information using the network itself: e.g. length of connections, other users' profiles visited and messages sent. SNSs provide a central repository accessible to a single provider. The high value of SNSs suggests that such data is being used to considerable financial gain.

- **Threat SN.3 Face recognition:** user-provided digital images are a very popular part of profiles on SNSs. The photograph is, in effect, a binary identifier for the user, enabling linking across profiles, e.g. a fully identified Bebo profile and a pseudo-anonymous dating profile.

- **Threat SN.4 CBIR:** Content-based Image Retrieval (CBIR) is an emerging technology which can match features, such as identifying aspects of a room (e.g. a painting) in very large databases, increasing the possibilities for locating users.

- **Threat SN.5 Linkability from image metadata:** many SNSs now allow users to tag images with metadata, such as links to SNS profiles (even if they are not the owner/controller of that profile), or even e-mail addresses. This leads to greater possibilities for unwanted linkage to personal data.

- **Threat SN.6 Difficulty of complete account deletion:** users wishing to delete accounts from SNSs find that it is almost impossible to remove secondary information linked to their profile such as public comments on other profiles.

- **Threat SN.7 SNS spam:** unsolicited messages propagated using SNSs. This is a growing phenomenon with several SNS-specific features.

- **Threat SN.8 Cross site scripting (XSS), viruses and worms:** SNSs are vulnerable to XSS attacks and threats due to 'widgets' produced by weakly verified third parties.

- **Threat SN.9 SN aggregators:** these 'SNS portals' integrate several SNSs which multiply vulnerabilities by giving read/write access to several SNS accounts using a single weak authentication.

- **Threat SN.10 Spear phishing using SNSs and SN-specific phishing:** highly targeted phishing attacks, facilitated by the self-created 'profiles' easily accessible on SNSs. SNSs are also vulnerable to social engineering techniques which exploit low entry thresholds to trust networks and to scripting attacks which allow the automated injection of phishing links.

- **Threat SN.11 Infiltration of networks:** some information is only available to a restricted group or network of friends, which should provide the first line of defence in protecting privacy on SNSs. However, since it is often easy to become someone's 'friend' under false pretences, this mechanism is not effective. On many SNSs it is even possible to use scripts to invite friends.

- **Threat SN.12 Profile-squatting and reputation slander through ID theft:** fake profiles are created in the name of well-known personalities or brands or within a particular network, such as a school class, in order to slander people or profit from their reputation.

- **Threat SN.13 Stalking:** cyberstalking is threatening behaviour in which a perpetrator repeatedly contacts a victim by electronic means such as e-mail, Instant Messenger and messaging on SNSs. Statistics suggest that stalking using SNSs is increasing.

- **Threat SN.14 Bullying:** SNSs can offer an array of tools which facilitate cyberbullying (i.e. repeated and purposeful acts of harm such as harassment, humiliation and secret sharing).

- **Threat SN.15 Corporate espionage:** social engineering attacks using SNSs are a growing and often underrated risk to corporate IT infrastructure.

Recommendations

The Virtual Group makes the following recommendations:

- **Recommendation SN.1 Encourage awareness-raising and educational campaigns:** as well as face-to-face awareness-raising campaigns on the sensible usage of SNSs, SNSs themselves should, where possible, use contextual information to educate people in 'real-time'. Additional awareness-raising campaigns should also be directed at software developers to encourage security-conscious development practices and corporate policy.

- **Recommendation SN.2 Review and reinterpret the regulatory framework:** SNSs present several scenarios which were not foreseen when current legislation (especially data protection law) was created. The regulatory framework governing SNSs should be reviewed and, where necessary, revised.

- **Recommendation SN.3 Increase transparency of data handling practices:** a review of the practices of SNS providers in Europe with respect to existing data protection law is recommended.

- **Recommendation SN.4 Discourage the banning of SNSs in schools:** SNSs should be used in a controlled and open way with co-ordinated campaigns to educate children, teachers and parents.

- **Recommendation SN.5 Promote stronger authentication and access-control where appropriate:** stronger authentication should be used in certain SNS environments. Additional authentication factors which could be used range from basic e-mail verification through CAPTCHAs[51] and recommendation-only networks to physical devices such as mobile phones and identity card readers.

- **Recommendation SN.6 Implement countermeasures against corporate espionage:** various steps are recommended for the prevention of social engineering attacks on enterprises.
- **Recommendation SN.7 Maximise possibilities for abuse reporting and detection:** SNSs should make it as easy as possible to report abuse and concerns. 'Report Abuse' buttons should be as ubiquitous as the 'Contact Us' option on classic websites.
- **Recommendation SN.8 Set appropriate defaults:** default settings should be made as safe as possible, and accompanied by user-friendly guidelines.
- **Recommendation SN.9 Providers should offer convenient means to delete data completely:** simple tools should be provided for removing accounts completely, as well as allowing users to edit their own posts on other people's public notes or comments areas.
- **Recommendation SN.10 Encourage the use of reputation techniques:** reputation mechanisms can act as a positive motivator towards good online behaviour.
- **Recommendation SN.11 Build in automated filters:** a legislative review into SNS filtering should be undertaken, with a view to SNS providers building filters into their sites.
- **Recommendation SN.12 Require consent from data subjects to include profile tags in images:** SNS operators should give users privacy tools to control the tagging of images depicting them.
- **Recommendation SN.13 Restrict spidering and bulk downloads:** SNS operators should restrict spidering and bulk downloads (except for academic research purposes).
- **Recommendation SN.14 Pay attention to search results:** data should either be anonymised, not displayed, or the user should be clearly informed that they will appear in search results and given the choice to opt out.
- **Recommendation SN.15 for addressing SNS spam:** similar techniques to those used for e-mail anti-spam reputation systems should also be developed to eliminate spam comments and traffic.
- **Recommendation SN.16 for addressing SNS Phishing:** the best practices for combating phishing on SNSs, which are promoted by the APWG, should be adopted.
- **Recommendation SN.17 Promote and research image-anonymisation techniques and best practices**
- **Recommendation SN.18 Promote portable Social Networks:** the economic and social implications of portable social networks should be addressed.
- **Recommendation SN.19 on research into emerging trends in SNS:** looking to the future, the group has identified some trends emerging in SNSs which have important security implications. More research should be carried out in the areas of mobile SNS, convergence with virtual worlds, misuse by criminal groups and 3D representation and online presence.

Introduction

Online Social Network Sites or Social Networking Sites (SNSs) are one of the most remarkable technological phenomena of the 21st century. User numbers have been increasing at a dramatic rate for several years. For example, as of June 2007, MySpace was the most visited website in the US with more than 114 million global visitors, representing a 72% increase on 2006. Facebook increased its global unique visitor numbers by 270% in the year ending June 2007 [1].

The defining characteristics of an SNS are:

- Tools for posting personal data into a person's 'profile' and user-created content linked to a person's interests and personal life
- Tools for personalised, socially-focused interactions, based around the profile (e.g. recommendations, discussion, blogging, organisation of offline social events, reports of events)
- Tools for defining social relationships which determine who has access to data available on SNSs and who can communicate with whom and how.

SNSs may be seen as informal but all-embracing identity management tools, defining access to user-created content via social relationships. The value of SNSs lies not just in the content provided (which is group-specific), but in its replication in electronic form of the web of human relationships and trust connections.

SNSs provide many benefits to their members:

- A sense of connectedness and intimacy (which is a healthy social enhancement), most often to an existing offline community but also to new online-only communities. There is evidence [2] that there is considerable social capital associated with the use of Facebook by US college students, which suggests that SNS use might contribute to increased self-esteem and satisfaction with life for some students.
- Tools which allow like-minded individuals to discover and interact with each other
- Identity-management and access-control tools for user-created content, allowing users to have control over who views their data (which is not generally permitted by blogs, for example)
- A forum for new modes of online collaboration, education, experience-sharing and trust-formation, such as the collection and exchange of reputation for businesses and individuals.

In addition to the benefits to members, SNSs have significant business value because of the marketing applications they offer. On SNSs, people profile themselves for free, and voluntarily disclose detailed maps of their social relationships. The figures speak for themselves: MySpace was sold in 2005 [3] for a price that corresponded approximately to 35 US$ per user profile. In 2006 Facebook sources suggested a valuation for their network of 2 billion US$ [4] (which would translate to 286 US$ per user profile) and, by September 2007, this figure had risen far higher [5] [6].

Since the success of an SNS depends on the number of users it attracts, there is pressure on SNS providers to encourage design and behaviour which increase the number of users and their connections. As with every fast-growing technology, however, security and privacy have not been the first priority in the development of SNSs. As a result, along with the above benefits, significant privacy and security risks have also emerged [7].

Users are often not aware of the size of the audience accessing their content. The sense of intimacy created by being among digital 'friends' often leads to inappropriate or damaging disclosures [8]. Social Networking may be seen as a 'digital cocktail party'. In general, the more contacts you have, the more popular you are, and the more influence you have. However, compared with a real-world cocktail party, SNS members broadcast information much more widely, either by choice or by mistake. For example, a brief survey of popular SNSs [9] shows several people openly publishing answers to 'surveys' with questions such as:

- Have you ever stolen money from a friend?
- Have you ever been in a fist fight?
- Have you ever cheated on a boyfriend/girlfriend?
- Have you ever drunk a bottle of alcohol by yourself?

Replies often appear in conjunction with a recognisable facial image of the person answering the survey. When combined with improvements in search technology [27] [28] this is likely to result in a significantly increased risk of incidents of personal damage. Some examples are given in [10] [14] [18].

The natural human desire to be connected with others, combined with the multiplying effect of SNS technology, can lead to a tendency to be inclusive rather than exclusive in accepting friend requests (i.e. lower the threshold for accepting friends). This is certainly not true of all users or all communities – certain more exclusive SNS communities have much higher average privacy thresholds. However, it is a prevailing driver and, since it tends to lead to faster network growth, it inevitably affects the sites with the largest number of users. This undermines the first line of defence for a user's data in SN security – the possibility of restricting access to a smaller network of contacts. It also contributes to the threat from viruses and worms spread via SNSs (see Threat SN.8).

Such possibilities, along with the threats posed by secondary data revealed to the service provider (see Threat SN.7), suggest that there may be a need to review current practice on SNSs with respect to data protection legislation and best practice, including the EU's 95/46 data protection directive [11], the OECD guidelines [12] and the US FTC's Fair Information Practices [13]. Conversely, since many of the trends emerging with SNS were not envisaged when these documents were drafted, there may also be a need for a review of best practice and legislation in the light of SNS scenarios.

The objective of this paper is to highlight and address privacy and security risks associated with SNSs. Such emphasis does not deny, discount or diminish the social, educational and economic value of SNSs. In fact, by highlighting possible risks and providing recommendations on how to minimise them, this paper offers strategies to improve privacy and security without compromising the benefits of information sharing – thus increasing the overall social value of SNSs.

Principal Threats

This chapter describes what, according to the contributors, are the most important privacy and security threats associated with SNSs. It focuses on threats which are specific to SNSs rather than those which are common to all web applications (e.g. identity theft, pharming, profiling), unless there is an SNS-specific variant (e.g. phishing, spam), or the threat is increased by some specific feature of SNSs (e.g. corporate espionage). Each threat description specifies the threat scenario and vulnerabilities (the technical or systemic weaknesses which lead to the risk) as well as the risks themselves (the potential negative impact).

Privacy Related Threats

Threat SN.1 Digital Dossier Aggregation

Vulnerabilities

Profiles on SNSs can be downloaded and stored over time and incrementally by third parties, creating a digital dossier of personal data. Information revealed on an SNS can be used for purposes and in contexts different from the ones the profile owner had considered.

Due to the greatly diminished costs associated with disk storage and Internet downloads, it is feasible to take regular snapshots of an entire network and store the profiles of its members indefinitely. On the other hand, it is costly and technically challenging to manage the complete deletion of data when it is no longer necessary. The information contained in individual profiles can be accumulated easily in order to track and highlight changes (e.g. list of girlfriends over time). This could be done either overtly or covertly by internal employees or external applications which can access profile information in bulk (e.g. through search features).

A common vulnerability is that more private attributes which are directly accessible by profile browsing can be accessed via search (e.g. a person's name and profile image is accessible via search on MySpace, Facebook and others, unless default privacy settings are changed).

Risks

Outside the social context of the network, information can become embarrassing or even damaging, as evidenced by reports of people missing out on employment opportunities due to employer reviews of SNS profiles [14] [15] [16]. While profiles can be changed or even deleted, additional storage elsewhere cannot be prevented – thus personal data takes on a life of its own even when the information itself may no longer be accurate or relevant. An adversary can retain information from SNS profiles for negative use in, for example, a blackmailing scheme or to embarrass the profile holder. In a recent case, the current Miss New Jersey was threatened with publication of images taken from her SNS profile if she would not give up her crown [17]. Recently, two tennis stars were suspended following revelations made on an SNS [18].

Threat SN.2 Secondary Data Collection

Vulnerabilities

In addition to personal data knowingly disclosed in a profile, an SNS member discloses personal information to the network operator using the network itself: data such as time and length of connections, location (IP address) of connection, other users' profiles visited, messages sent and received and so forth. While this in itself is not specific to SNSs, in other contexts information on a user's behaviour is protected to some degree by being spread across multiple websites, e-mail accounts (e.g. work e-mail and one or multiple private e-mail accounts), one or more instant messaging systems and multiple Internet access points (at work, at home, through a cell phone etc.).

Since SNSs are not currently portable (there is no widely used standard for exchanging SNS data and there is a significant overhead in joining a new network: defining relationships, profiles, inviting contacts and so forth), there is a strong tendency towards amalgamating all SNS activity under a single provider. This is then a powerful data warehouse for the owners of the SNS. Despite this, there is currently a lack of transparency about certain data collection practices. For example, it is not made clear to users how and to whom the data about visits to other profiles are made public. Privacy policies tend to be vague in specifying what is and what is not personal information.

The following is an example of a privacy statement:

"[SNS Provider] also logs non-personally-identifiable information including IP address, profile information, aggregate user data, and browser type, from users and visitors to the site. This data is used to manage the website, track usage and improve the website services. This non-personally-identifiable information may be shared with third-parties to provide more relevant services and advertisements to members." [19]

This does not specify which elements of 'profile information' are disclosed to third parties. While some elements of a person's profile, such as their hobbies, are arguably not classified as personally identifiable, the high value paid for SNSs (e.g. $35 per user for MySpace in 2006) strongly suggests that the data they contain is being used to considerable financial gain. Referring to the Article 29 Working Party's 2007 Opinion on the concept of personal data [20], the use of profile data in the personalisation of advertising should certainly be regarded as personal data.

Risks

Due to the expansion of SNSs from a mere collection of profiles to a one-stop communication hub for messaging services, interest groups, video content and more, network operators and application providers can now gather unprecedented amounts of secondary personal information on their users. While information disclosed is ostensibly used by the network operator to customise and personalise its services, it can also be used for targeting (e.g. advertising), discrimination (e.g. price discrimination) or the transfer of data to third parties through resale.

Threat SN.3 Face Recognition

Vulnerabilities

User-provided digital images are an integral and exceedingly popular part of profiles on SNSs. As an example, Facebook hosts in excess of 1.7 billion user photos (as of 21 May 2007), a database growing at a rate of more than 60 million per week [21] [22]. Since images are tied to individual profiles and often either explicitly (through, for example, labelled boxes on images) or implicitly (through recurrence) identify the profile holder, they constitute a data source suitable for correlating profiles across services using face recognition.

The efficiency of face recognition algorithms has improved dramatically over the last decade [23] [24]. While early systems performed well only on carefully controlled images [25], newer systems can handle a wide variety of image conditions [26]. The combination of better algorithms with faster computing hardware and essentially unlimited storage enables comparisons of large numbers of images [26].

Until recently, the use of face recognition software has been the domain of law enforcement and border-control agencies. Recently, however, some mainstream web service providers have announced work on integrating face recognition technologies into their applications [27] [28].

Both CBIR (see Threat SN.4) and face recognition are part of a broader threat posed by so-called 'mashups', which link data between independently provided web services to provide previously unforeseen inferences including highly personal information.

Risks

If successful, the usage of face recognition allows the linking of image instances (and the accompanying information) across services and websites. This enables connecting, for example, a fully identified Bebo profile with a (potentially intentionally different) pseudo-anonymous Friendster profile or a pseudo-anonymous dating profile with an identified corporate website profile. The photograph is then, in effect, a binary pseudonym for the user, which can be linked across profiles in the same way as a traditional pseudonym. As a result, an adversary can gather substantially more information about a user than intended. While any network user can

already establish these connections manually by simply looking at the profiles, the large number of profiles renders any organised manual effort of linking profiles unfeasible. The risk associated with manual re-identification has been demonstrated many times in the context of college life pranks [10].

Threat SN.4 CBIR (Content-based Image Retrieval)

Vulnerabilities

Related to face recognition, Content-based Image Retrieval (CBIR) [29] was originally developed for digital forensics. CBIR is an emerging technology which is able to match features, such as identifying aspects of a room (e.g. a painting) in very large databases of images. Traditional search terms are replaced with a reference image or image template. Search is designed to be resilient to cropping, resizing, rotation and quality adjustment (e.g. for JPEG) [30] [31] [32] [33] [34]. Currently, privacy controls on images uploaded and the advice given on SNSs do not take into account the possibility of CBIR, and few people are aware of the consequences of posting images with location-specific content online.

Risks

While face recognition allows the linking of profile data involving the person's physical body, CBIR allows linking of location data through the recognition of common objects in images. CBIR opens up the possibility of deducing location data from apparently anonymous profiles containing images of users' homes. This can lead to stalking, unwanted marketing, blackmail and all the other threats associated with unwanted disclosure of location data. It can also assist blackmailers looking for specific types of image which might later be used as part of a digital dossier (Section 1). Not only can it help them to find compromising material, by targeting specific profiles for download, it could also help them to circumvent 'spider throttling' mechanisms which limit the number of successive page-loads in a specific time-window from a given IP address.

Threat SN.5 Linkability from Image Metadata, Tagging and Cross-profile Images

Vulnerabilities

Many SNSs now allow users to tag images with metadata such as the name of the person in the photo, a link to their SNS profile (even if they are not the owner/controller of that profile), or even their e-mail address. As an example, shown in the following extract from the Facebook help pages, Facebook allows tagging of images with profile data and even e-mail addresses.

> *"Can I tag people who do not use Facebook in photos?*
>
> *You can tag whomever you want. While tagging your photos, if you type in the name of someone who is not on your Friend List, you have the option of listing their e-mail address. When you are done tagging, they will receive an e-mail that provides a link to the image. They will be allowed to see the photos in which they are tagged, but nothing else on the site unless they register."* [35]

Even if users exercise caution over which images they post of themselves and their location, their privacy may be under even greater threat from images posted by others. While profile links can usually only be included for profiles in a person's friend list, given the low trust threshold for inclusion in this list, this does not offer much consolation. Very few SNSs offer privacy tools to control the tagging of images with links to their profile or the accessibility of tagged images in search results.

Another aspect of image metadata is that many cameras embed metadata about the camera in the image including, in many cases, the serial number of the camera. Given that many cameras are linked to address data through warranty registration cards, this constitutes a threat to the user's privacy. An interesting recent case was the posting of a full illegal copy of *Harry Potter and the Deathly Hallows* which included embedded versions of the serial number of the camera used to take it, as well as the exact date and time the images were taken [36].

Threat SN.6 Difficulty of Complete Account Deletion

Vulnerabilities

Users wishing to delete accounts from SNSs will find that, although it is usually very easy to remove their primary pages, secondary information such as public comments they have made on other accounts using their identity will remain online. Moreover, in general there is ambiguity as to whether information will be deleted upon account closure. As an example, the Facebook privacy policy makes the statement:

> *"Removed information may persist in backup copies for a **reasonable** period of time but will not be **generally** available to members of Facebook."* [37]

Upon 'deactivating' an account, users of some providers such as Facebook receive an e-mail telling them how to 'reactivate' their account – implying that a copy is kept of personal data. Furthermore, personal data cannot be completely deleted unless users manually remove all public notes or comments on other profiles. This is usually not feasible due to the large number of steps involved [38].

Risks

The user loses control over his/her identity. Damaging comments cannot be removed, increasing the 'digital dossier' effect. Users cannot exercise their fundamental right to control over their own personal information. This means that sites which do not provide easy means for deleting or rectifying information may be in contravention of the European Privacy Directive 95/46, which states:

> *"Every reasonable step must be taken to ensure that data which are inaccurate or incomplete, having regard to the purposes for which they were collected or for which they are further processed, are erased or rectified". Data should be "kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected or for which they are further processed."*
>
> *"Member States shall guarantee every data subject the right to obtain from the controller … as appropriate the rectification, erasure or blocking of data … because of the incomplete or inaccurate nature of the data."* [11]

SNS Variants of Traditional Network and Information Security Threats

Threat SN.7 SN Spam

Vulnerabilities

SN spam is unsolicited messages propagated using SNSs [39]. Many spammers have sought to capitalise on the exponential growth of SNSs and the free traffic they provide. This is a very serious issue since statistics suggest that SNSs are replacing e-mail in some circles as a method of communication. This means that the same scale of spam problems which have affected e-mail communications systems could soon affect SNSs.

Common techniques used by spammers include:

- The use of specialised SNS spamming software such as FriendBot [40] to automate friend invitations and note/comment posting. Such tools use the SNSs' search tools to target a certain demographic segment of the users and communicate with them from an account disguised as that of a real person.
- The sending of notes typically including embedded links to pornographic or other product sites designed to sell something.
- Friend invitations, using an attractive profile which is likely to persuade someone to accept the invitation. The profile or the invitation then contains links to external sites advertising products or even phishing for passwords.
- The posting of spam comments on public notes or comments areas of 'friends'. Typically, spammers will create as many 'friends' as possible, focussing on those with public notes or comments areas or message boards and fitting a specific demographic profile (Friendbot has features to automate this) and then post spam messages on their public notes or comments areas.
- Stealing members' passwords to insert and promote their offers on another profile.

Until recently no filters were available for notes or friend requests. The best a user could do was block notes from the sender's address. MySpace, for example, now includes an option for users to 'report spam/abuse' addresses. Spammers, however, frequently change their address from one throw-away account to another.

The following is a typical spam friend request:

"Hey Everyone! I've moved my profile here because MySpace won't allow me to post some of my nude modelling pictures. If you want to see my more revealing pictures, I've uploaded my entire modelling photo album to my free profile here. Click Here For My Personal Pictures And Video [Link to spam site] which allows some of my more scandalous photos. Signing up takes 2 seconds as they just want to verify you are 18 or over. After you've signed up simply search for my handle "Sexy4U2" to get to my 'personal' page". [41]

Already, SNS providers have implemented filters that attempt to slow the use of robots, and are aggressively deleting 'spam accounts' when discovered.

Risks

The risks are broadly the same as with other kinds of spam, i.e.:

- Traffic overload
- Loss of trust or difficulty in using the underlying application
- Phishing and diversion to pornographic sites
- One risk specific to SNSs is that, because profiles are created specifically for spamming, and Sybil attacks (a type of reputation attack) on spam protectors involve the creation of large numbers of false profiles, the SNS can become 'diluted' by fake profiles which reduce its value to legitimate users.

Threat SN.8 Cross Site Scripting, Viruses and Worms

Vulnerabilities

In some SNSs, users can post HTML within their own profiles and message-boards. SNSs are particularly vulnerable to XSS (cross site scripting) attacks. So-called 'widgets', produced by weakly verified third parties, are widely used [42]. In addition, a heavy reliance on message-posting and viral marketing means SNS viruses spread extremely quickly. The SAMY virus [43], which infected MySpace profiles, had spread to over one million users within just 20 hours, making it one of the fastest spreading viruses of all time [44] [45]. This is part of a wider problem; very short development cycles, driven by sharp increases in numbers have led to a neglect of secure development practices.

Risks

Among the effects of such vulnerabilities are:

- Account compromise
- Denial of service (SAMY forced MySpace to shut down its site) and associated loss of reputation
- Diversion to phishing attacks
- Unsolicited content may also be spread to e-mail and IM traffic (SNSs generate e-mail and IM messages).

Threat SN.9 SNS Aggregators

Vulnerabilities

Social Aggregators such as Snag, ProfileLinker and many others [46] are relatively new applications, which address the problem of having to set up social networks on multiple platforms by integrating data from various SNSs into a single web application. This unfortunately multiplies the vulnerability of accounts by giving read/write access to several SNS accounts based on a single weak username/password authentication. Social Aggregators also carry the risk of increasing the potential for attackers to mine data across sites since many offer cross-SNS search features.

Risks

Many of the risks are the same as with any kind of authentication compromise. They include:

- Identity theft
- Zombification of SNS accounts, e.g. for XSS attacks or advertising
- Loss of privacy for other members of the SNS by allowing search across a broader base of data.

Identity-related Threats

Threat SN.10 Spear Phishing using SNSs and SN-specific Phishing

Vulnerabilities

Spear phishing describes any highly targeted phishing attack. The existence of easily accessible self-created 'profiles' and self-declared 'circles of friends' on SNSs allows a phisher to harvest

large amounts of reliable social network information which may be used for a highly personalised phishing attack. An experiment conducted by researchers at the University of Indiana showed that, using data available on SNSs, e-mail phishing attacks can achieve a hit rate of 72%, compared with a control of 15% [47].

A related threat is the use of SNSs for the phishing attack itself (rather than just gathering data to be used elsewhere). The worm JS/Quickspace.A was designed to spread through MySpace profile pages. Pages were infected with links to a phishing site which then asked for the user's logon details and used these to embed a link to the phishing site in the stolen profile [48]. Although this is merely a new modality of an existing threat, the extra trust created by the 'circle of friends' can make this a particularly effective form of phishing attack.

The spread of such phishing attacks is greatly increased by the vulnerability of SNSs to social engineering techniques based on the infiltration of SNSs with low entry thresholds. Another factor is the prevalence of scripting attacks allowing the automated injection of phishing links.

Risks

Spear phishing using SNSs carries more or less the same risks as other kinds of phishing, including:

- Compromised logins (e.g. of the SNS profile) which can in turn increase the speed of spread of a phishing attack
- Identity theft
- Financial damage
- Reputation damage.

Threat SN.11 Infiltration of Networks Leading to Information Leakage

Vulnerabilities

Some information is only available to 'friends' or members of a restricted group and this is the first line of defence in protecting privacy on SNSs. Since it is often very easy to become someone's 'friend' under false pretences, this mechanism is not very effective. Currently it is even possible to use scripts to invite friends on MySpace, and the growing amount of specialised commercial software such as Friendbot [49] and FriendBlasterPro [50], created for exactly this purpose, suggests that this is increasingly common. CAPTCHAs [51] are not implemented by default and in fact a significant number of 'SNS spam' invitations do get sent (see Threat SN.7). Some SNSs have extremely broad criteria for membership of a network (and access to data within that network). For example, currently anyone with an appropriate e-mail address can join any geographical (i.e. city) network on the Facebook and gain access to the public profiles of this network.

Social and commercial pressure to get as many friends as possible means there is often a tendency to accept friend requests without checking their authenticity or suitability. In a recent experiment, antivirus company Sophos created a profile page for 'Freddi Staur' (an anagram of 'ID Fraudster'), a green plastic frog with only minimal personal information in his profile. They then sent out 200 friend requests to see how many people would respond, and how much personal information could be gleaned from the respondents. The following are some of the results:

- 87 of the 200 users contacted responded to Freddi, with 82 leaking personal information (41% of those approached)
- 72% of respondents divulged one or more e-mail address
- 84% of respondents listed their full date of birth [52].

Risks

While this does not cause much direct damage other than polluting SNSs with irrelevant or misleading profiles and thereby reducing their utility, it is an enabler for many of the other threats discussed in this paper. For example, it allows attackers to:

- View private information
- Phish for information and even physical contacts using the trust gained
- Conduct spamming and marketing campaigns

Threat SN.12 Profile-squatting and Reputation Slander through ID Theft

Vulnerabilities

Fake profiles are created in the name of well-known personalities or brands or in order to slander people who are well known within a particular network of friends (e.g. a school class). Not all profiles should be an accurate portrayal of the individual posting the profile; profiles exist of many dead celebrities, which may have a serious educational value both for readers and profile-writers. For example Galileo has a profile on MySpace (as well as over 3000 friends) [53].

However, when fake profiles are used for malicious purposes such as defamation, serious damage can be done. While this is also possible using conventional web pages, SNSs provide an extra dimension because:

- The connection to an SNS makes it easier to target the abuse at the people who are most likely to notice it (e.g. a teacher's class).
- The main purpose of SNS profiles is to describe the person they purport to represent. It is therefore generally assumed that there will be a single profile for each real individual and it is more likely to be assumed that profile information has been created by the individual it represents.
- The target of the attack may not be able to access the profile – it may be restricted to the group which is ridiculing the person (as in the case of the pupils of a teacher). It is very unlikely that the real person behind the profile is part of the SN.
- Most SNSs perform only weak authentication of registrants (weaker than domain registration, for example, where at least a credit card must usually be provided).

Risks

The risks are similar to domain squatting (registration of domain names of well-known brands, with high prices to release them).

Profile-squatting can lead to:

- Libel and personal damage. Fake profiles are used to damage reputation or ridicule someone in public, simply out of revenge, or in order to blackmail them [54]. This is not just an issue for celebrities. For example there have been a number of incidents reported of profiles being used to ridicule teachers or fellow pupils in schools [55].
- Phishing. Such profiles can be used to lure unsuspecting users into divulging information which can be used for phishing attacks. For example, by masquerading as a person's offline friend (which is not difficult to do), it is possible to trick them into giving out location data.
- Marketing under false pretences. False profiles are used to advertise products while pretending to be a 'friend' of the target.
- Legal action against perpetrators with no malicious motives. Setting up fake profiles may be seen as a fun activity which may even have educational value.

Social Threats

Threat SN.13 Stalking

Vulnerabilities

Stalking typically involves threatening behaviour in which the perpetrator repeatedly seeks contact with a victim through physical proximity and/or phone calls (offline stalking), but also by electronic means such as e-mail, Instant Messenger and messaging on SNSs (online or cyberstalking). There are not many reliable statistics on stalking, but those that are available suggest that stalking using SNSs is increasing [56].

SNSs encourage the publication of personal information, including data that can reveal an individual's location and schedule (for instance, home address and home phone, schedule of classes and so on) or a person's online usage (for example, an instant messaging profile that can reveal the online status of the user). In a 2005 study of one university's Facebook network, between 15 and 21% of users disclosed both their full current address as well as at least two classes they were attending. Since a student's life is mostly dominated by class attendance, the combination of address and class schedule provides the physical location of the user throughout most of the day (and night). A much larger number of users, 78%, provided instant messaging (IM) contact information suitable for tracking their online status [57]. Emerging mobile-based SNSs such as Twitter [58] tend to emphasise location data even more [59]. It can also be seen from the other threat descriptions that SNS

provides many other more subtle methods for stalkers to track their targets.

Risks

The impact of cyberstalking on the victim is well known and can range from mild intimidation and loss of privacy to serious physical harm and psychological damage.

Threat SN.14 Bullying

Vulnerabilities

Cyberbullying is a term used to describe repeated and purposeful acts of harm that are carried out using technology, particularly mobile phones and the Internet. Research in the area is in its infancy and, in common with quantitative research into other forms of abuse, statistics vary from study to study and should only be regarded as indicative. What is apparent is that reported instances of cyberbullying via SNSs are increasing [60] [61]. A 2006 study [62] found that:

> *"About one out of ten youngsters have been involved in frequent cyberbullying: 3.3% exclusively as a victim, 5.0% exclusively as a perpetrator, and 2.6% as both a victim and a perpetrator."*
>
> *"The majority of youngsters (63.8%) believe cyberbullying is a 'big problem'. This figure may reflect either a general assessment of the issue in the eyes of the youngsters, or it may indicate that they find it a serious problem for those being bullied."*

Whether this is due in whole, in part or in combination to the increased use and development of SNSs, increased platform compatibility, increased access to the Internet, ease of multimedia creation and distribution, or indeed to the increasing recognition that there are a group of acts which utilise technology that are identifiable as bullying is not currently known. SNSs tend to offer an array of tools to users – for example, in addition to profile and people search there may also be blogging or micro-blogging facilities, instant messaging, chat rooms, community and collaboration areas etc. which together constitute a very useful 'suite' of tools for the bully. Each of these elements can be used positively or potentially misused.

A number of factors make SNSs particularly vulnerable to this kind of exploitation:

- Many schools ban the use of SNSs at school, which acts as a strong disincentive to the reporting of bullying. A US National School Boards Association survey [63] claims 52% of schools ban the use of SNSs on campus.
- The ease of remaining anonymous (using a fake profile).
- The ease of communicating with restricted groups of people (a feature which can be very beneficial if used for the right purposes).
- The one-stop-shop effect. SNS provides all the usual tools and attacks used by a cyberbully, and more, in a single interface (IM, mobile messaging, fake profiles and slander, controlled broadcast of slanderous messages)[64].
- The generation gap. Teachers and adults are frequently unable to intervene because they are not familiar with the technology used.

Risks

In *Cyberbullying and Cyberthreats*[65], Willard identifies several forms of cyberbullying behaviour that can be carried out on SNSs (although none of them are exclusive to SNSs and many have obvious offline comparisons) (see overleaf).

Forms of Cyberbullying Behaviour that can be carried out on SNSs [65]

Flaming: Online fights using electronic messages with angry and vulgar language.

Harassment: For example, repeatedly sending hurtful or cruel and insulting messages; gaining access to another's username and password in order to send inappropriate messages to friends' lists.

Denigration: Setting up accounts pretending to be people in order to humiliate them; sending or posting gossip or rumours about a person to damage his or her reputation or friendships, e.g., the creation of 'Hate' websites, the posting of jokes, cartoons, gossip and rumours, all directed at a specific victim; posting harmful, untrue and/or cruel statements or pictures, and inviting others to do the same, or to comment on them.

Impersonation: Pretending to be someone else and sending or posting material to get that person in trouble, put them in danger or to damage their reputation or friendships.

Outing: Sharing someone's secrets or embarrassing information or images online.

Trickery: Talking someone into revealing secrets or embarrassing information, then sharing it online.

Exclusion: Intentionally and cruelly excluding someone from an online group, for example, a group of offline friends deciding to ignore a specific individual as a form of punishment.

Stalking: Typically linked to a problematic intimate relationship, repeated, intense harassment and denigration that includes threats or creates significant fear.

Threatening behaviour: Either direct or indirect (interestingly, Willard includes threats to hurt someone or to harm oneself).

Threat SN.15 Corporate Espionage

Vulnerabilities

Social engineering attacks using SNSs are a growing and often underrated risk to corporate IT infrastructure. Social engineering is a means of attack frequently used by hackers to bypass security mechanisms and access sensitive enterprise data – not by using technology (although technology may be involved), but by using the employees themselves. Data is often acquired subtly and is gathered gradually piece by piece.

SNSs can be a particularly important tool in an organised social-engineering attack on an enterprise. Some information is necessary to enter an online community but often the privacy settings are neglected and therefore the threshold for gaining information to be used in a social engineering attack is very low. For example, several professional SNSs publish information on lists of employees [66]. For example, an SNS search results page lists employees currently or previously working at Barclays Bank, which could be useful to someone collecting information for a social engineering attack on an enterprise. This vulnerability is specific to SNSs since it allows attackers to see the connections between employees.

There could also be information about stakeholders with whom a company is doing business on an SNS. If an employee publishes sensitive information (for example, position, qualification and/or function) on an SNS, this might pose a serious threat to a company [67]. The publication of information about IT infrastructures, such as network directories, is often seen on technical blogging or discussion forums, but also on SNSs.

Risks

The main risk here is the loss of corporate intellectual property, but gaining access to insiders may also be a component in a broad range of other crimes, such as hacking corporate networks to cause damage, blackmailing of employees to reveal sensitive customer information and even to access physical assets.

Government Policy Recommendations

Rec. SN.1 Encourage awareness-raising and Educational Campaigns

Threats: All, but especially Digital Dossier, Face Recognition, Cyberbullying, SNS Spam, Corporate Espionage

As well as face-to-face awareness-raising campaigns, SNSs themselves should, where possible, use contextual information to educate people in 'real-time'.

This already happens in some cases, but it should be encouraged as best practice. Sites should also publish user-friendly community guidelines rather than 'terms and conditions'; these are much less intimidating to users. Accessible language should be used so that users can easily understand the rules of the site – the clearer the guidelines, the more likely users are to abide by them.

- Profiles may be captured continuously and stored and caches exist even for data which is apparently deleted.
- A person may be recognised by images, especially of their face, on the Internet, just as in the offline world.
- The size or nature of the audience which has access to content may not be as expected in an offline circle of friends.
- Accepting untrusted friend requests can lead to spam and phishing.
- Images contain information which can be used to pinpoint location or identify a person.
- Images can also give away private data about other people especially when tagged with metadata.
- Images may contain embedded data which identify the device used to shoot them and thereby indirectly identify the owner.
- Profile information may appear in some search results even if you believe it to be private.
- The potential abuses of information posted on SNSs (e.g. when it contains location data) for the purpose of stalking. While many SNSs already restrict users from posting location data, it is virtually impossible to prevent users from posting it unwittingly in messages and posts on public notes or comments areas.

Many SNSs already ban certain data types (e.g. zip codes). Best practice as to banned data types on SNSs should be defined and promoted on all sites. At a minimum, users should be encouraged to refrain from public disclosure of real-world contact information (e.g. home address and fixed or cell-phone number).

Typical advice given to minors to address cyberbullying is [68]:

- Tell a trusted adult about the bullying – and keep telling until the adult takes action.
- Do not open or read messages from cyberbullies.
- Tell your school if it is school-related. Schools have bullying policies in place.
- Do not erase the messages – they may be needed to take action.
- Never agree to meet with the person or with anyone you meet online.
- If you are threatened with harm, inform the local police or ask your parents to do so.
- Consult the increasing amount of educational material provided on safe usage of SNSs (some produced by providers themselves) [69] [70] [71] [72].
- Parents should look carefully at any pictures their children are posting on these sites. Is there more information in the picture than was intended, such as hobbies, interests or the location of their school?

Additional awareness-raising campaigns should also be directed at software developers to encourage security conscious development practices and corporate policy.

This is often ignored in the rush to roll out the next iteration, particularly given the phenomenally rapid expansion of SNSs.

Rec. SN.2 Review and Reinterpret Regulatory Framework

Threats: Secondary Data Collection, Digital Dossier, Face Recognition, CBIR, Squatting

SNSs present several scenarios which were not foreseen when current legislation (especially data protection law) was created. This means that certain issues may need to be clarified. In some cases, the existing legal framework may even need to be modified or extended.

The regulatory framework governing SNSs should be reviewed and, where necessary, revised.

Specific issues include:

- What is the legal position on deletion of user-generated content by service providers if it is classed as SNS spam?
- What is the legal position on image-tagging by third parties?
- Who is responsible for security flaws resulting from user-generated markup or scripting?
- How should privacy policies of embedded third party widgets be communicated to users?
- What exactly constitutes personal data in an SNS environment?
- What is the legal position on profile-squatting?
- Should the posting of certain classes of data by minors (location data) be made illegal?

Rec. SN.3 Increase Transparency of Data-handling Practices

Threats: Secondary Data Collection, Digital Dossier, Account Deletion

A review of practices of SNS providers with respect to existing data protection law and best practice is recommended (e.g. [11] [12] [13]). European data protection law, for example, requires clear and explicit notice to be given to data subjects of:

- The purpose for which the data is used (including secondary usage)
- Any third party recipients of the data
- The existence of a means of access and rectification.

The transparency and accuracy of data handling statements, especially relating to third party widgets including mood indicators and survey responses from identified individuals, should be examined since current language is often vague and uninformative.

Users should be given accurate information on what is done with their data before and after account closure.

Examples of questions requiring greater transparency

- What is done with data on profile visits?
- What data is transmitted to widget providers?
- How can a privacy policy be accessed for third party widgets?
- What are the secondary purposes of the processing of profile data?
- A clear description of the difference between the 'deactivation' and the 'closure' of an SNS profile. In particular, how long is a profile kept after it is deactivated?

Descriptions of practices should be conveyed in a user-friendly way, with important information being conveyed in the context in which it is relevant, rather than being buried in Terms and Conditions.

Rec. SN.4 Discourage the Banning of SNSs in Schools

Threats: Stalking, Bullying

A growing number of schools are banning or restricting the use of SNSs in schools [73]. It is recommended that schools and education policy-makers should carefully consider the consequences of banning SNSs since this acts as a disincentive to the reporting of bullying. It also means that teachers and adults are less likely to learn the skills needed to mentor and monitor young people in this area. Finally it also means that a valuable educational resource is lost.

SNSs should be used in a controlled and open way (i.e. not banned or discouraged), with co-ordinated campaigns to educate children, teachers and parents.

This would have a wider knock-on effect as many of the vulnerabilities described in this paper can be addressed simply by raising awareness and as

children in turn educate their parents and teachers.

It is not the technologies themselves which are responsible for bullying behaviour but the individuals who misuse them. For this reason, education, the modelling of the positive use of technology by peers, teachers and adults and community self-regulation are all key areas in combating cyberbullying.

Provider and Corporate Policy Recommendations

Rec. SN.5 Promote Stronger Authentication and Access-control where appropriate
Threats: Digital Dossier, Ease of Infiltration, Squatting, Spear Phishing, Stalking, SNS Spam, Social Aggregators

The strength of authentication chosen in SNS environments depends on the SNS. Many SNSs benefit from the ability to masquerade as another persona, and therefore this is not appropriate. The benefits can include educational profiles, safety through anonymity and harmless experimentation. Virtual worlds are in fact a manifestation of this type of network. However, in certain types of SNS, both users and providers can benefit from stronger authentication and the greater validity this lends to claims made on SNSs. Examples include so-called white-label SNSs such as Ecademy [74] and more professional networks, such as LinkedIn, which are used as a basis for business contacts. On all networks, authentication methods which can differentiate bona fide members from spammers are also useful.

It may also be possible that, if the process of stronger authentication were made more user-friendly, it would not act as a disincentive to enrolment. In fact it could have the opposite effect by increasing the trust placed in others on the network.

There are a number of additional authentication factors that SNSs could use to enhance their offer by reducing the level of fake and troublesome memberships. These range from basic e-mail verification through CAPTCHAs and recommendation-only networks (where enrolment is only through invitation) to physical devices such as mobile phones and identity card readers, where these have been deployed (e.g. as part of a wider eID pilot [75]). Each method has its own characteristics in terms of ease of use, strength of authentication and additional privacy risk (e.g. some users may feel it is very intrusive to have their profile linked to a government-issued identity number; while others may not wish to divulge their mobile phone numbers), so there is unlikely there will be a single solution.

Rec. SN.6 Implement Countermeasures against Corporate Espionage using SNSs
Threats: Corporate Espionage

A key factor in the prevention of social engineering attacks is employee security awareness. Unfortunately, there is no silver bullet for recognising an attack before it leads to major damage. Every company has its own characteristics and vulnerabilities. Just as a social hacker has to familiarise himself with these peculiarities, protective measures have to be tailored to a company's specific requirements in order to be effective. Protection can only be achieved by enabling staff to recognise the difference between defined processes and requests that deviate from these definitions. This, of course, requires in-depth security awareness among employees.

The following steps are recommended for the prevention of social engineering attacks caused by information leaked on SNSs:

- Make employees aware that they need to be as careful online as in real life.
- Establish a security policy including the use of SNSs.
- Promote the idea that the more information is given out, the more vulnerable you are.
- At a minimum, SNS providers should require membership of a network before revealing its members or their relationships. This is particularly important since no single member of the network (e.g. Barclays Bank) is responsible for such a privacy setting.

Rec. SN.7 Maximise Possibilities for Reporting and Detecting Abuse

Threats: Bullying, Spear Phishing, Infiltration, Profile Squatting, SNS Spam, Cross Site Scripting

Systems and policies for handling illegal action and activity that breaks terms and conditions should be built into the design of the application. For instance, if a member of the public reports an offensive image on a certain group or page there should be a well-documented procedure for how this will be dealt with, incorporating protection against bogus reports and, where appropriate, using reputation aggregation systems to make judgements. Similar systems and policies should be in place for law-enforcement related concerns.

SNSs should make it as easy as possible to report abuse and concerns. 'Report Abuse' buttons should be as ubiquitous as 'Contact Us' options on classic websites

Rec. SN.8 Set Appropriate Defaults

Threats: All

Few users change default settings [76] therefore it is vital that these are made as safe as possible. Defaults could also be tailored to the (claimed) age of the person signing up, since it may be appropriate to set different default privacy settings for minors than for over 21s.

A set of guidelines should be produced on appropriate default settings which could encourage the use of best practices by application-providers. Furthermore, (proposed) portable SNS formats (see Section 18) should include privacy preferences so that users are not discouraged from selecting stricter settings simply by the set-up time required when moving to a new application.

Default settings should be made as safe as possible, and accompanied by user-friendly guidelines.

Rec. SN.9 Providers should offer Convenient Means to Delete Data Completely

Threats: Account deletion

Simple, easy to use tools should be provided for removing accounts completely, as well as allowing users to edit their own posts on other people's public notes or comments areas. Privacy policies and help pages should explain clearly how this can be done.

Technical Recommendations

Rec. SN.10 Encourage the use of Reputation Techniques

Threats: Ease of Infiltration, Squatting,Stalking, Cyberbullying, SNS Spam, Cross Site Scripting

Reputation-based techniques can be a very effective means of determining the authenticity of users and their claims about themselves, preventing many of the threats detailed above.

Possible uses of reputation techniques in SNSs include:

- Filtering of malicious or spam comments
- Filtering comments by quality to increase content quality
- Increasing reliability of third party widgets
- Reporting inappropriate or copyrighted content
- Reporting profile-squatting or identity theft
- Recommendation-only sign-up (where new members have to be introduced by an existing member) e.g. [77]. This requires a good balance between setting entry hurdles too high and viral (but weakly authenticated) growth.
- Reporting of inappropriate behaviour and posting of high-risk data such as location information.

This practice relies on the collective goodwill of the majority but experience shows that, given the opportunity, most users will help to protect the safety and quality of a project. It also leads to a more engaged community, which is willing to take more responsibility for security rather than relying on the service provider.

Reputation mechanisms can also act as a positive motivator towards good online behaviour, since most people enjoy the effects of a good reputation and respond to appropriate feedback. This is clearly demonstrated, for example, in collaborative online content applications such as Slashdot [78] – human moderation does not work best when hidden from view and smart interface design can significantly reduce moderation time (and cost).

For more information, see ENISA's Position Paper, 'Reputation-based Systems: a security analysis'[79].

Rec. SN.11 Build in Automated Filters

Threats: SNS Spam, Spear Phishing, Cross Site Scripting and Viruses

The first line of defence against offensive, litigious or illegal content should be smart filters. Filters should not replace human intervention – they will never understand slang trends or cultural sensitivities, for example – but they can remove content such as a single user who is making automated submissions. If there is a particular piece of content that is driving significant volumes of traffic, automated filtering tools can often determine if it is legitimate content. They can also be highly effective when combined with reputation systems [80].

In the area of e-mail filtering, it was necessary to clarify legislation in order to allow service providers to delete spam messages without fear of legal consequences. There may also be a need for a legislative review in the area of SNS filtering.

Rec. SN.12 Require the Consent of the Data Subject to Include Profile Tags or e-Mail Address Tags in Images

Threats: Image metadata

SNS operators should give users privacy tools to control the tagging of images depicting them.

The tagging of images with personal data without the consent of the subject of the image violates the user's right to informational self-determination (the control over who publishes their data and where). SNS operators should implement mechanisms for giving users control over who tags images depicting them. This could be done by including a setting in each profile with the following options:

- Allow anyone to tag any image with this profile
- Request the profile owner's consent before tagging
- Do not allow tagging.

Obtaining consent could be facilitated by, for example, an automatic e-mail with a consent button which is sent out as soon as a tagging operation is requested. Inclusion of an e-mail address should always trigger a consent request of this type.

Rec. SN.13 Restrict Spidering and Bulk Downloads

Threats: Digital Dossier, CBIR

SNS operators should restrict spidering and bulk downloads (except for academic research purposes).

SNS operators should protect all means to access profiles which might lend themselves to bulk access. They should also put in place access restrictions that make it harder to create bogus accounts. Many operators such as Facebook are already operating policies which restrict the bulk download of content. This is in the interests of both users and service providers since it protects the competitive advantage providers have by

having a large data set for personalising their own services. It also protects users from later blackmail and lack of control over their own data. Specifically, SNS operators should protect all means to access profiles (e.g. search, poke [81]) etc. which might lend themselves to bulk access. SNS operators should also put in place access restrictions that make it harder to create bogus accounts. Measures such as CAPTCHAs and bandwidth throttling are a good first line of defence here. However, exceptional allowance should be made to access data sets for the purposes of academic research.

Rec. SN.14 Provide more Privacy Control over Search Results

Threats: Digital Dossier, Stalking, SNS Spam, Corporate Espionage, SNS Aggregators

SNS providers should take care not to allow data to appear in search results when users believe it is private.

> Data should either be anonymised, not displayed, or the user should be clearly informed that it will appear in search results and given the choice to opt out.

Search results often give access to data which is otherwise restricted, thereby giving data miners a powerful tool to aggregate private information. One privacy policy states:

> *"Your name, network names, and profile picture thumbnail will be available in search results across the Facebook network and those limited pieces of information may be made available to third party search engines"* [82]

– i.e. search results appear to have a different status from directly discovered profiles. This feature exists for the obvious reason that, without it, it would be very difficult to discover new users (it would have to occur via an introduction of a third party). LinkedIn allows search by organisation, in many cases, allowing the creation of an instant employee directory. The ability to tag photos with metadata about other people's profiles means that a cross-profile search for images of a particular person is now possible.

Rec. SN.15 Recommendations for Addressing SNS Spam

Threats: SNS Spam

SNSs have several advantages over e-mail when it comes to detecting spam. Profile comments are a key indicator because it is very difficult to get fake friends to post real discussions – no comments suggests a fake account. A spammer would need to create a few dozen profiles and replicate the thread of discussion via their profiles, so that it could make someone's profile look 'real'.

However, what looks real to a human being and what looks real to a software algorithm are often different. Keyword analysis is not enough. Tools should be developed in a similar vein to link doping [83] and spamdexing [84] defences which would use metrics based on the topology of social networks to detect spammers.

Rec. SN.16 Recommendations for Addressing SNS Phishing

Threats: Spear Phishing and SNS-specific Phishing

As well as awareness-raising, the best practices for combating phishing are already promoted by the APWG [85]. For example:

- Flag or even ban links which do not point to the text shown to the user
- Flag or ban links on images representing text links (using OCR).

Research and Standardisation Recommendations

Rec. SN.17 Promote and Research Image-Anonymisation Techniques and Best Practices

Threats: Face Recognition, CBIR

Users should be aware that an image is a binary pseudonym which can help in linking profiles across sites. While algorithms for face de-identification are currently under development [7], they are not widely available yet. Such algorithms can limit the ability of automatic face recognition software to recognise and link faces by removing identifying information while preserving other aspects of the face such as gender, ethnicity and expression. Users can limit the ability of automatic face recognition software to recognise and link their faces by avoiding usage of identical images across services as well as choosing images which are difficult for algorithms to recognise, i.e. non-frontal images under non-standard illumination, displaying varied facial expressions. Network operators, some of which already actively encourage users to upload face images, could give similar recommendations and check for compliance.

Further research should be done into obfuscation tools which can make images more difficult to recognise by automated tools or even tools for making facial images unrecognisable to human viewers. Many tools exist which allow morphing of images to make them into caricatures of the original, but these have been shown to be reversible [86]. A range of transformations should be available from an avatar representing the user without looking like them at all to minor modifications which may render them unrecognisable to a machine but still recognisable to a human viewer.

Rec. SN.18 Promote Portable Networks

Threats: SNS Aggregators, Secondary Data Collection, Digital Dossier

Many of the threats outlined above, in particular those relating to data privacy, have arisen because current SNSs are extremely centralised (i.e. high numbers of users with few providers). Where users were previously protected by spreading their data over many mutually inaccessible repositories, it is now collected in a single place. It is currently very difficult to transfer your social network from one provider to another, or to interact between providers. The result is that people tend to gravitate towards the most popular providers so that they do not have to invite their friends repeatedly which has a high time and social overhead. Currently, for every new SNS community, site users have to re-enter all personal profile information (name, e-mail, birthday, URL etc.), friends and privacy preferences. This also adds another barrier to the safe use of available privacy features since time spent on setting up features is wasted in transporting them to another site. Another related trend has therefore been towards SNS aggregators which integrate existing providers using ad hoc data interfaces.

While there are clear commercial reasons behind these trends, the security and usability implications of a centralised and closed data storage model should not be ignored. A possible solution to this problem is portable social networks, which allow users to control and syndicate their own ‘social graph’ along with privacy preferences, blocked users and filter settings [87]. At a minimum, it should be possible to export the social graph and its preferences from one provider to another and, ideally, users would have the possibility of complete control over their own social data, syndicating it to providers which created added-value ‘mashup’ applications. Some proposals already exist for such formats [88] but the economic and social implications also need to be addressed in order for them to succeed.

Rec. SN.19 Research into Emerging Trends in SNSs

Threats: All

Looking to the future, the group has identified some trends emerging in SNSs which have important security implications.

More research should be carried out in the areas of:

- Mobile SNS
- Convergence with virtual worlds and 3D representation
- Misuse by criminal groups
- Online presence

Mobile SNS: There is a trend towards increased usage of mobile-based SNSs such as Twitter. As an illustration, Twitter has grown its user-base to 122,000 monthly unique visitors within the space of just over one year [89]. The increase in location information disclosed has important security implications [90].

Convergence with virtual worlds and 3D representation: Given the similar aspirations of many SNS and virtual world users and the extra functionality offered by virtual worlds, a widespread convergence between the two seems only a matter of time. In fact applications such as Kaneva [91] are already offering this kind of convergence and Google is reported to be testing a '3D Social Networking application' [92]. Such applications may introduce new security threats such as those related to virtual world economics [93] [94].

Misuse by criminal groups: While there are many benefits of a tool which allows like-minded individuals to discover and interact with each other, this feature could also be abused, in that it allows for collaboration on illegal activity. Research should be conducted into the ways and the extent to which SNSs are used by criminal groups.

Online presence: Increasing amounts of information and tools are available relating to online presence (whether someone is currently online and logged into a particular site), or even physical location, which is typically revealed more in mobile-based SNSs like Twitter. More research is needed into the privacy and security implications of online presence.

Concluding Remarks

The take-home message of this paper is that SNSs have clear benefits to society, not least because they herald the end of passive media, bringing free interactive user-generated content to anyone with an Internet connection. Social Networking is fundamentally an Identity Management system. If used correctly, it can enhance data privacy over and above more established mechanisms such as blogs. If not, however, it provides a dangerously powerful tool in the hands of spammers, unscrupulous marketers and others who may take criminal advantage of users. New technologies such as online face-recognition tools, combined with the false sense of intimacy often created by SNSs, can lead to a serious erosion of personal and even physical privacy.

User-generated content should be accompanied by attention to security and privacy issues in the development of code and data-handling policies. Most importantly, users should be educated in how to use social media safely via awareness-raising on the sites themselves and in schools – targeted at students, parents and teachers. This would also address the increasing danger of a 'digital divide' between those with the know-how to join in the social-software revolution and those without. It requires a culture-shift in educators from the 'there be dragons' scaremongering attitude of banning (or trying to ban) SNS usage to a more mature attitude of encouraging sensible, well-informed use.

Finally, this is a matter for Governments as well as service providers and end-users. Legislation and policy is currently not equipped to deal with many of the challenges that social media present. Education policy should reflect the urgent need to educate both young and older users, students, teachers and parents on how to benefit from SNSs without suffering their downsides. Legislation should be reviewed and interpreted to fit the new paradigms with which we are faced.

References and Links

1 Comscore media metrix, July 2007
www.comscore.com/press/release.asp?press=1555

2 N. Ellison, C. Steinfield, C. Lampe. *The Benefits of Facebook Friends: Social Capital and College Students' Use of Online Social Network Sites*

3 BBC News, 19 July, 2005. *News Corp in $580m Internet buy*
http://news.bbc.co.uk/2/hi/business/4695495.stm

4 Ian Sefferman. *By the Numbers – Is Facebook worth $2 Billion?*, 2006
www.iseff.com/2006/04/by-numbers-is-facebook-worth-2-billion.html

5 K. Delaney, R. Guth, V. Vara. *Microsoft Fires Volley At Google in Ad Battle*, 2007
http://online.wsj.com/article/SB119065193646437586.html?mod=hpp_us_whats_news

6 K. Swisher. *How High Can you Count?: New Facebook Fundraising*, 2007
http://kara.allthingsd.com/20070911/how-high-can-you-count-new-facebook-fundraising/

7 R. Gross and L. Sweeney. *Towards real-world face de-identification*, IEEE Conference on Biometrics: Theory, Applications and Systems, 2007

8 A. Acquisti and R. Gross. *Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook*, Workshop on Privacy Enhancing Technologies (PET), 2006.
http://privacy.cs.cmu.edu/dataprivacy/projects/facebook/facebook2.pdf

9 Retreived from MySpace forum posting, September 2007
http://forum.myspace.com/index.cfm?fuseaction=messageboard.viewThread&entryID=426692&groupID=100096072&adTopicID=23&Mytoken=D01E608F-7C6B-4FF9-A62F3F0A3E469281844594

10 K. Maternowski. Campus police use Facebook, *The Badger Herald*, 25 January 2006
http://badgerherald.com/news/2006/01/25/campus_police_use_fa.php

11 Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data, Official Journal of the European Communities of 23 November 1995 No L. 281 p. 31
http://ec.europa.eu/justice_home/fsj/privacy/docs/95-46-ce/dir1995-46_part1_en.pdf

12 OECD Guidelines on the Protection of Privacy and Transborder Flows of Personal Data
http://www.oecd.org/document/18/0,2340,en_2649_34255_1815186_1_1_1_1,00.html

13 US Federal Trade Commission, *Fair Information Practices In The Electronic Marketplace*, 2000
www.ftc.gov/reports/privacy2000/privacy2000.pdf

14 A. Fuller. Employers snoop on Facebook, *The Stanford Daily*, 20 January 2006
http://daily.stanford.edu/article/2006/1/20/employersSnoopOnFacebook

15 J. Flesher. Wall Street Journal Career Site, *How to Clean Up Your Digital Dirt Before It Trashes Your Job Search*, 2006
www.careerjournal.com/jobhunting/usingnet/20060112-flesher.html

References and Links

16 S. Gilbertson. Delete your bad web reputation, *Wired*, 2006
www.wired.com/science/discoveries/news/2006/11/72063

17 E. Pilkington. *Blackmail claim stirs fears over Facebook*
www.guardian.co.uk/international/story/0,,2127084,00.html

18 BBC News, *Tennis LTA suspends top junior players*, 2007
http://news.bbc.co.uk/sport2/hi/tennis/7010983.stm

19 Facebook Privacy Policy
www.facebook.com/policy.php

20 Article 29 Data Protection Working Party, Opinion 4/2007 on the concept of personal data
http://ec.europa.eu/justice_home/fsj/privacy/docs/wpdocs/2007/wp136_en.pdf

21 D. Beaver. *Facebook photos infrastructure*, 2007
http://blog.facebook.com/blog.php?blog_id=company&m=5&y=2007

22 Facebook blog on image database size
http://blog.facebook.com/blog.php?post=2406207130

23 R. Gross, J. Shi and J. Cohn. *Quo vadis face recognition?*, Third Workshop on Empirical Evaluation Methods in Computer Vision, 2001

24 W. Zhao, R. Chellappa, P. J. Phillips and A. Rosenfeld. *Face recognition: a literature survey*, ACM Computing Surveys, 2003

25 P. J. Phillips, P. Grother, R. J. Michaels, D. M. Blackburn, E. Tabassi and J. M. Bone. FRVT 2002: Evaluation Report, March 2003

26 P. J. Phillips, W. T. Scruggs, A. O'Toole, P. Flynn, K. Bowyer, C. Schott and M. Sharpe. FRVT 2006 and ICE 2006 Large-Scale Results, NISTIR 7408, March 2007

27 Jacqui Cheng. *Facial recognition slipped into Google's image search*, Ars Technica, 2007
http://arstechnica.com/news.ars/post/20070530-facial-recognition-slipped-into-google-image-search.html

28 Example CBIR providers
www.polarrose.com/, www.myheritage.com

29 Content Based Image Recognition, Wikipedia entry
http://en.wikipedia.org/wiki/CBIR

30 CBIR Demonstration, Washington University Computer Science Department
www.cs.washington.edu/research/imagedatabase/demo/

31 G. Richard. *Digital Forensics*, Presentation
www.cs.utexas.edu/%7Eshmat/courses/cs395t_fall05/richard.ppt

References and Links

[32] Chen, Y., Roussev, V., Richard, G. III, & Gao, Y. (2005). Content-based image retrieval for digital forensics, in *Proceedings of the First International Conference on Digital Forensics (IFIP)*, 2005
www.cs.uno.edu/%7Egolden/Stuff/IFIP2004.pdf

[33] M. Sutter, T. Müller,R. Stotzka et al. *Inspector Computer*, German eScience Conference, 2007
www.aip.de/groups/escience/GeS2007/contributions/GES_paper50.pdf

[34] R. Datta, D. Joshi, J. Li, and J. Z. Wang. *Image Retrieval: Ideas, Influences, and Trends of the New Age*, ACM Computing Surveys, 2007 (ACM-CSUR 2007)
http://infolab.stanford.edu/%7Ewangz/project/imsearch/review/JOUR/datta.pdf

[35] Retrieved from Facebook help pages (September 2007), tagging images
www.facebook.com/help.php?page=7

[36] S. Schoen. Harry Potter and Digital Fingerprints, *Deep Links News*, 2007
www.eff.org/deeplinks/archives/005371.php

[37] Retrieved from Facebook Privacy Policy
www.facebook.com/policy.php

[38] Siyavash. *Corporate Facebook*, 2007
http://siyavash2005.googlepages.com/facebook

[39] Wikipedia entry for Social Networking, Spam
http://en.wikipedia.org/wiki/Social_networking_spam

[40] Friendbot, automated friend adding software
www.friendbot.com/

[41] Spam message received on MySpace profile, August 2007

[42] E. Skoudis. *What are the risks of SNSs?*
http://searchsecurity.techtarget.com/expert/KnowledgebaseAnswer/0,289625,sid14_gci1247616,00.html

[43] SAMY virus description, Wikipedia
http://en.wikipedia.org/wiki/Samy_(XSS)

[44] Timeline of notable computer viruses and worms, Wikipedia
http://en.wikipedia.org/wiki/Timeline_of_notable_computer_viruses_and_worms

[45] J. Grossman, Whitehat Security. *Cross-Site Scripting Worms And Viruses – The Impending Threat And The Best Defense*, 2006
www.whitehatsec.com/downloads/WHXSSThreats.pdf

[46] List of Social Network Aggregators
http://mashable.com/2007/07/17/social-network-aggregators/

[47] T. Jagatic, N. Johnson, M. Jakobsson, F. Menczer. *Social Phishing*, ACM, October 2007
www.indiana.edu/~phishing/social-network-experiment/phishing-preprint.pdf

References and Links

48 Worm description JS/Quickspace.A
www.f-secure.com/v-descs/js_quickspace_a.shtml

49 Friendbot, automated friend adding software
www.friendbot.com/

50 Friendblasterpro, friend adding software
www.addnewfriends.com/

51 CAPTCHA – Completely Automated Public Turing test to tell Computers and Humans Apart
http://en.wikipedia.org/wiki/Captcha

52 Sophos website, *Sophos Facebook ID probe shows 41% of users happy to reveal all to potential identity thieves*, August 2007
www.sophos.com/pressoffice/news/articles/2007/08/facebook.html

53 MySpace profile of Galileo
http://profile.myspace.com/index.cfm?fuseaction=user.viewprofile&friendid=93676521

54 *Fake Celebrity Pages on MySpace*, Digital Info, 2006
www.mydigitallife.info/2006/07/15/fake-celebrity-pages-on-myspace/

55 Associate Press, *PA Principal Sues Students over Fake MySpace Profiles*, 2007
www.foxnews.com/story/0,2933,264622,00.html

56 2006 Cyberstalking statistics (Note increase in SN initiated stalking from 2005)
www.haltabuse.org/resources/stats/2006Statistics.pdf

57 R. Gross and A. Acquisti. Information revelation and privacy in social networking sites, Workshop on Privacy in the Electronic Society (WPES), 2005

58 Twitter – Mobile-based SN
http://twitter.com/

59 Mashup showing location of twitter messages
http://twittermap.com/maps

60 I-safe advice against cyberbullying
www.isafe.org/channels/sub.php?ch=op&sub_id=media_cyber_bullying

61 *Putting U in the picture – Mobile bullying survey 2005*, NCH and Tesco Mobile, 2005
www.nch.org.uk/uploads/documents/Mobile_bullying_%20report.pdf

62 Ed Robby Deboelpaep. *European Parliamentary Technology Assessment, Cyberbullying among Youngsters in Flanders*, 2006
www.viwta.be/files/executive%20overview%20cyberbullying.pdf

63 *Creating & Connecting Research and Guidelines on Online Social – and Educational – Networking*, US National School Boards Association, August 2007
www.nsba.org/site/docs/41400/41340.pdf

References and Links

64 Netsafe, advice on reporting bullying and identity theft on SNSs
www.netsafe.org.nz/youngadults/bullying_social_networking.aspx

65 N. Willard. *Cyberbullying and Cyberthreats, Responding to the Challenge of Online Social Aggression, Threats, and Distress*, Research Press, 2007

66 Example of a list of employees available via LinkedIn
www.linkedin.com/search?search=&sortCriteria=3&company=%22Barclays+Bank+plc

67 M. Sunner, MessageLabs. *Deadly Sins: Emerging Technologies, Threats & Trappings*, Presentation at ISSE 2007
www.eema.org/downloads/isse2007/presentations/sumner.pdf

68 i-SAFE Internet Safety Tips for Parents
www.isafe.org/verizon/docs/tipsheets.pdf

69 Bebo online safety video
www.bebo.com/Safety.jsp?MID=4428327486

70 *Social Networking Sites: Safety Tips for Tweens and Teens*, US Federal Trade Commission
www.ftc.gov/bcp/edu/pubs/consumer/tech/tec14.pdf

71 Online Guard, Social Networking Sites, a parents' guide (also in Spanish)
http://onguardonline.gov/socialnetworking.html

72 Safefamilies, Safety Tips for Social Networking
www.safefamilies.org/socialnetworking.php

73 A. Van Duyn. How skills are slipping through the net, *Financial Times*, 27 September 2007
www.ft.com/cms/s/1/093a2938-6d15-11dc-ab19-0000779fd2ac.html

74 Ecademy – white label SN
www.ecademy.com/

75 *Towards an eID Pilot*, documentation for EU Large Scale eID pilot project
http://ec.europa.eu/information_society/activities/ict_psp/library/call_docs/docs/eid_pilot.pdf

76 W. Mackay. *Triggers and barriers to customizing software*, proceedings of CHI'91, pages 153–160, ACM Press, 1991

77 A recommendation-only SN
www.asmallworld.net/

78 Slashdot reputation-based moderation system
http://slashdot.org/moderation.shtml

79 ENISA Position Paper on 'Reputation-based Systems: a security analysis'
www.enisa.europa.eu/doc/pdf/deliverables/enisa_pp_reputation_systems.pdf

80 List of collaborative content filtering research papers
http://jamesthornton.com/cf/

References and Links

81 Facebook Poke feature
www.facebook.com/help.php?page=20

82 Facebook Privacy Policy
www.facebook.com/policy.php

83 Wikipedia definition for Link Doping
http://en.wikipedia.org/wiki/Link_doping

84 Wikipedia definition for Spamdexing
http://en.wikipedia.org/wiki/Link_spam

85 Anti Phishing Working Group
www.antiphishing.org/

86 Germany's federal police force reverses photo morphs found on the Internet, which had been digitally altered to disguise a suspected paedophile's face.
www.interpol.int/Public/THB/vico/Default.asp

87 Microformats Wiki, Social Network Portability
http://microformats.org/wiki/social-network-portability

88 FOAF project
www.foaf-project.org/

89 Siteanalytics monthly people count for Twitter.com
http://siteanalytics.compete.com/twitter.com

90 Twitter vision – mashup showing location of twitter messages
http://twittervision.com/

91 Kaneva, 3D social networking application
http://www.kaneva.com/

92 J. Cheng Google testing 'My World' for launch later this year
http://arstechnica.com/news.ars/post/20070924-google-testing-my-world-for-launch-later-this-year.html

93 E. Feldman. *Real Crime in Virtual Worlds*, Associated Content, 2007
www.associatedcontent.com/article/329951/real_crime_in_virtual_worlds.html

94 Reuters, *Virtual currencies need real-world laws to prevent crime*, 2007
www.itpro.co.uk/security/news/112850/virtual-currencies-need-realworld-laws-to-prevent-crime.html

For further information about this Position Paper,
contact Giles Hogben
(giles.hogben@enisa.europa.eu)

ENISA – European Network and Information Security Agency
PO Box 1309, 710 01, Heraklion, Crete, Greece
Tel: +30 2810 39 12 80, Fax: +30 2801 39 14 10
www.enisa.europa.eu

03/12_07