ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

Security and Privacy in Big Data

T.Shanmuga Vadivu[1], D.Saranya[2], S.Karthika[2]

Assistant Professor, Dept. Computer Science, Sri Krishna Arts & Science College, Coimbatore, India[1]

II-M.Sc(CS) Student, Dept. Computer Science, Sri Krishna Arts & Science College, Coimbatore, India[2]

Abstract: Although the use of Social Networking web sites and applications is increasingly on the rise, many users are not properly informed of the risks associated with using these sites and application. Understanding these risks and challenges should be addressed to avoid potential loss of private and personal information. Current authentication systems suffer from many weaknesses. Many available graphical passwords have a password space that is less than or equal to the textual password space. . In this paper, we present and evaluate our contribution, i.e., the 3-D password. The 3-D password is a multifactor authentication scheme. This paper examines the issues of security, privacy, and trust in social networking sites from users' viewpoint.

Keywords: 3D password, Security, authentication, privacy, and trust in social networking sites.

I. INTRODUCTION

The rapid growth of SNS in recent years indicates that they are now a mainstream technology for many people. According to BobIvins, vice president of comScore.com, “social networking is not a fad but rather an activity that is being woven into the very fabric of the global Internet.” A variety of social networking sites (SNSs) are used by hundreds of million users. The people who use social networking sites see them as a fun and easy leisure activity. Through SNS, users can keep in touch with friends and family, especially with people they do not see on a regular basis, find old friends, contact friends of friends, and even contact people they haven’t met before. By extending their social circle, users have the opportunity to communicate with people who have the same interests.

Users provide personal information about themselves including their interests, social relationships, current occupation, pictures and other media content, share this information via SNSs Platforms Due to the sensitivity of information stored within social networking sites a plethora of research in the area of information security has been conducted.

However, since the reputation of these SNS has been tarnished by a number of incidents in news media, such as the massive worldwide spam campaign in Quechup, sexual predators, stalkers, child molesters...about their privacy. The dramatic increase of computer usage has given rise to many security concerns. One major security concern is Authentication, which is the process of validating who you are to whom you claimed to be. In general, human authentication techniques can be classified as knowledge based (what you know), token based (what you have), and biometrics (what you are). While there is a continual flow of media stories discussing privacy and security problems of SNSs, the great majority of academic contributions focus either exclusively on possible threats on one hand, or possible protection strategies on the other. When it comes to privacy and security issues on social networks, "the sites most likelys to suffer from issues are the most popular ones," Graham Cluley, Chief Technology Officer at UK tech security firm Sophos says. But security issues and privacy issues are entirely two different beasts. A security issue occurs when a hacker gains unauthorized access to a site's protected coding or written language. Privacy issues, those involving the unwarranted access of private information, don't necessarily have to involve security breaches. Someone can gain access to confidential information by simply watching you type your password. But both types of breaches are often intertwined on social networks, especially since anyone who breaches a site's security network opens the door to easy access to private information belonging to any user. But the potential harm to an individual user really boils down to how much a user engages in a social networking site, as well as the amount of information they're willing to share. In other words, the Facebook user with 900 friends and 60 group memberships is a lot more likely to be harmed by a breach than someone who barely uses the site.

The reason social network security and privacy lapses exist results simply from the astronomical amounts of information the sites process each and every day that end up making it that much easier to exploit a single flaw in the system. Features that invite user participation -- messages, invitations, photos, open platform applications, etc. -- are often the avenues used to gain access to private information, especially in the case of Facebook. In response to the potential threats that users are exposing to, most of the major networks now enable users to set privacy controls for who has the ability to view their information. In this paper Section II outlines current security and privacy threats regarding social networking sites, section III gives the possible protection mechanism for SNS. Section IV is the privacy framework which gives blueprint for secured SNS. Section V deals with the implementation of above mentioned section. The aim of this paper is thus to provide an introduction to both state-of-the-art attack scenarios as well as possible mitigation

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

strategies for social networking sites, to ultimately spot potential gaps between attacks and defenses.

II. ISSUES

The top and foremost privacy problem is that SNS do not inform users of the dangers of divulging their personal information.

2.1 Privacy related threats

a) Digital dossier aggregation

SNS profiles can be fetched and stored by third parties in order to create a digital dossier of personal data. Hogben et al argue that due to diminished costs of disk storage and Internet downloads it is feasible to take incremental snapshots of entire SNSs. A proof-of concept digital dossier aggregation, carried out on an early version of the most popular German SNS (meinVZ), showed that 1.074.574 profiles could be aggregated within less than four hours with a computer cluster consisting of ten computers highlighted various methods how data could be collected from Facebook. A commercial provider even offers packages for crawling Social networks which can be used to aggregate publicly available information.

b) Secondary data collection vulnerabilities

SNS members also disclose information to their Internet service providers (ISPs). While this is not solely limited to SNSs, the main difference is the extent of coherent personal data exposed to ISPs. For example, to map the circle of friends without SNSs data, ISPs need to correlate information from multiple Email addresses, instant messaging, etc. Even more important is the threat of disclosure and resale of personal information to third parties, for example to providers of targeted advertisement. At the time of writing no case of secondary data collection has been documented. A recent case with AT&T however illustrated how serious this threat is.

c) Face recognition vulnerabilities

SNS users provide profile images of themselves and SNSs contain shared images associated with them. Face recognition technology can be used to identify users across different SNSs, no matter if pseudonyms or fake names are being used.

d) CBIR (Content-based Image Retrieval)

CBIR is a technology which deduces the location of users by analyzing and comparing common patterns in images. Hence shared images within SNSs not only disclose the identity of users but possibly the location of users as well.

e) Linkability from Image Metadata, Tagging and Cross-profile Images

While users control which information and media they share within a SNS, they can't control which content other users upload and link to their profile. Images might also contain metadata including the serial number of the camera used to make the pictures.

f) Difficulty of Complete Account Deletion

Users that wish to deactivate their SNS account face difficulties to do so in most cases. On the one hand because not all comments and messages sent to other users will be deleted, and on the other hand because SNS providers keep backups of account data. Most social networking sites offer the possibility to permanently delete a user account, this features are however often hidden from users. In the case of Facebook users have to follow a special link which can only be found through a search within the Facebook support center.

2.2 Trolling

A common misuse of social networking sites such as Face book is that it is occasionally used to emotionally abuse individuals. Such actions are often referred to as trolling. It is not rare for confrontations in the real world to be translated online. Trolling can occur in many different forms, such as (but not limited to) defacement of deceased person(s) tribute pages, name calling, playing online pranks on volatile individuals and controversial comments with the intention to cause anger and cause arguments. Trolling is not to be confused with **cyber-bullying.**

2.3 Online bullying

Online bullying, also called **cyber-bullying**, is a relatively common occurrence and it can often result in emotional trauma for the victim. Depending on the networking outlet, up to 39% of users admit to being "cyber-bullied". There are not many limitations as
to what individuals can post when online. Individuals are given the power to post offensive remarks or pictures that could potentially cause a great amount of emotional pain for another individual.

2.4 Cyber-bullying and grooming

Cyber-bullying are aggressive attacks and bullying attempts carried out over the Internet, while cyber-grooming refers to attempts by adults to approach minors via the web to abuse them sexually. One of the most infamous cases involving cyber-bullying, the "Megan Meier case", led to the suicide of a teenage girl. In the Meg Meier case the perpetrator exploited the ease of setting up a fake profile, which was also used in a recent cyber-grooming case.

2.5 Stalking

SNSs can be misused by perpetrators to contact their victims but also to gather information on them. SNSs users often disclose location data via their pictures or personal information.

2.6 Cross Site Scripting, Viruses and Worms

In order that users are able to customize the design of their profiles, SNSs often provide the possibility to post HTML

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

code. Furthermore third party applications (widgets) are used to extend the functionality of SNSs and together with HTML code they state a risk for Cross-site scripting (XSS) vulnerabilities. Samy/JS.Spacehero for example was a XSS worm on MySpace, which infected more than one million profiles within the first 24 hours. A number of worms targeted other social networking sites like Face book, MySpace, and Orkut.

2.7 Third Party Threats

Users also have no control over third parties. Users cannot add an application to their profile without granting it permission to access all their public and private data. A simple application Send a Rose which allows a user to send roses to his friends requires unnecessary full access to the user's data. This increases the risk of having "Attractive" applications that spy on users and collect their data. Face book additionally gives third parties second-degree access.

2.8 Profile Cloning

The technique of stealing social network user's identity is called profile cloning. The main targets of profile cloning are users who set their profiles to be public. Public profile allows attackers to obtain profile information easily, and therefore can duplicate or copy their profile information to create a false identity. There are two types of profile cloning.

2.9 Existing Profile Cloning

In existing profile cloning, attackers create a profile of already-existing users by using their name, personal information, as well as picture to increase reliance, and then sending friend requests to friends of that user. This action is successful since most users accept friend requests from the person that they already know without looking through it carefully. Also, it is possible that a person might have multiple accounts. If victims accept the friend requests, then attackers will be able to access their information.

2.10 Cross-Site Profile Cloning

In cross-site profile cloning, attackers steal user's profile from one social networking site that users register an account, and then create a new user's profile on another social networking site that user has not registered on before. After that, attackers use users contact list from the registered social networking site to send a friend requests to all those contacts in another social networking site. In this case, it is more convincing than the first case since there is only one account for that particular user. Then, if the contacts accept friend request, attackers can access their profile

III. PROTECTION STRATEGIES

Recently various data protection schemes have been proposed to protect the user's privacy in social networks. Common methods for defense include the use of encryption, data dissociation or the usage of fake information. A combination of these methods is likely to protect the users' privacy to a larger extend.

Encryption can be used to secure communication channels. In the most naive approach this means that the communication between the users and the social network uses encryption (e.g., HTTPS) to protect against eavesdropping. However, this from a technical standpoint simple, easy applicable and readily available protection instrument is not widely used by most of the SNSs. XING is the exception, as it uses HTTPS for all client communication. Encryption can be further used to protect content distribution like user to- user communication on the SNS without modifications to the underlying infrastructure. This defeats an honest, but curious SNS operator from eavesdropping, as well as an adversary that is able to get access to the data in any unauthorized way.

Fake information can be used as an additional layer of protection against curious social networking operators or external adversaries. The social network only sees the fake information, while possibly authentic and sensitive information is stored encrypted on a third party server. As a source for fake information either predefined wordlists or dynamic content from the Internet might be used. This prevents naive approaches for detecting the fake information. NOYB for example shuffles user data among all NOYB users to increase privacy, based on a cryptographic pseudorandom algorithm. It is implemented as Firefox extension and uses a public dictionary of all users as input. This means that the NOYB users can hide among all NOYB users, while it still remains relatively hard to detect and works without changing the underlying infrastructure of Facebook. The Firefox extension Face Cloak on the other hand uses a slightly different approach, by using random articles from Wikipedia as source for fake data and custom wordlists as source for fake names. This provides strong privacy against Facebook and unauthorized users. Despite the fake information, encrypted real information is stored on a third party server.

Stenography might be used to embed information in pictures or videos hosted or exchanged over SNS. As the videos and pictures are transformed upon submission to fit the size constraint of the websites, the steganographic algorithms need to be robust enough to withstand these transformations. NOYB for example relies on stenography as one possible communication channel.

3-D password is a multifactor authentication scheme. It can combine all existing authentication schemes into a single 3-D virtual environment. This 3-D virtual environment contains several objects or items with which the user can interact. The type of interaction varies from one item to another. The 3-D password is constructed by

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

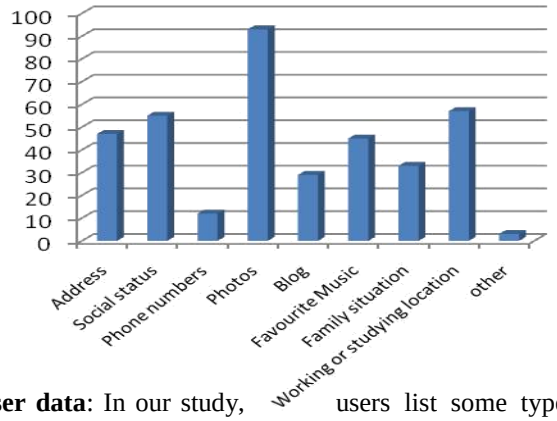
Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

observing the actions and interactions of the user and by observing the sequences of such actions. It is the user's choice to select which type of authentication techniques will be part of their 3-D password.

IV. PRIVACY FRAMEWORK

The role of the Privacy Framework is to provide a foundation for SNS in which privacy issues can be addressed. In this part, we categorize user data, user privacy concerns and profile viewers. Based on these categorizations, we present four privacy levels (No Privacy, Soft Privacy, Hard Privacy, Full Privacy) and three tracking levels (Strong Tracking, Weak Tracking and No Tracking).



User data: In our study, users list some type of information that they would place on their profile. Based on the survey, we categorize user data into 5 groups: Identity, Demographic profile, Activity, Social Network, and Added content.

Identity refers to information that makes it possible to determine physically who the user is. This includes information such as name, address or telephone number.

Demographic refers to the demographic characteristics of the user, such as age, gender, weight, race, and/ or political view. **Activity** lists all the activities that users perform within the SNS, for instance: adding new Friends, writing a comment in profile of other users, and/ or changing their status. The Activity data is automatically collected by the SNS provider and is displayed in News Feed format.

Social Network refers to the relationships of users in SNS, such as who are their Friends or the groups they subscribed to.

Added content is all the information that users put on their profile page, including blog, photos, music or video clips.

User privacy concern

Since different users have different privacy concerns for each piece of information, we propose four **Privacy settings** for user data according to impact on user privacy: Healthy, Harmless, Harmful and Poisonous.

Healthy data is general information about users such as nick name, usual hobbies, landscape photos, and music video clips. Specifically, if an unauthorized person accesses this data, it cannot be tracked back to the user. The user can confidently share this data without any privacy concern.

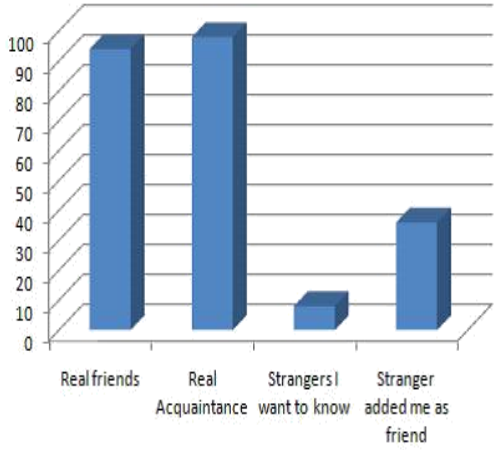
Harmless data contains the user's demographic profile, such as gender, religion, age groups, and political views. Specifically, the disclosure of harmless data does not create either Security risks orReputation and Credibility risks. However, it can lead to Profiling since some marketing companies can collect this data and build a profile of the user.

Harmful data refers to inappropriate photos or blog entry that may damage the user's reputation, for example a photo of Alex in his job uniform smoking pot. This data can damage the Reputation and Credibility of the user.

Poisonous data contains information that may cause Security risks such as the user's financial information, name, address, SIN (Social Insurance Number)... Cyber criminals can use this data for identity theft purposes.

Profile Viewers

These four Privacy settings: Healthy, Harmless, Harmful and Poisonous indicate to what extent the information disclosure can cause privacy risk to the user. Nonetheless, this categorization of user data is not sufficient by itself. Specifically, the level of security threats depends, not only on the type of data being disclosed, but also on the person to whom it is being disclosed. In our study, most of users' online friends are real friends and real acquaintances. However, some of them (24.8%) are ready to become friends with strangers also.



Thus, according to the intimacy and trust among users in SNS, we classify people who can see the user profile into

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

four basic groups: Best Friends, Normal Friends, Casual Friends, and Visitors.

Best Friends are people that the user trusts enough to share nearly everything with. They often are best friends of the user in real life.

Normal Friends can be the user's family members, relatives or friends in real life. **Casual Friends** usually are people about whom the user only knows a little. The user may only be acquainted with them online. **Visitors** could be users or non-users of the SNS. They usually can only see the user's nickname or his avatar. They are not in the Friend list but they may be able to see user's avatar or some personal information such as name, age and location.

Privacy levels

Based on these four basic groups, we adapted the four levels of privacy in to the context of SNS

No Privacy the user does not care about the privacy of his personal information. Everyone can see all his information on the SNS

Soft privacy: the user wants to keep his Poisonous data only for Best Friends. The Visitors are allowed to see Harmless and Healthy data of the user. The Casual and Normal Friends can access to all user data, except the Poisonous one.

Hard privacy: The Normal Friends still can access to Harmful data but the user put more limit on Visitors as they can only see the Healthy data and the Casual Friends only have access to Harmless and Healthy data.

Full privacy: the user does not allow Visitors to access his data. The Poisonous and Harmful data are restricted to Best Friends and the Normal and Casual Friends can access Harmless and Healthy data only.

Tracking levels

Besides privacy levels, the user also worries about being tracked through profiles of other users on SNS. There are three possible ways of tracking a user on SNS: following a profile link in a Friend list of a user, follow a name tag of a user, and reading information about a user in one of his Friends' profile.

Strong tracking The user does not mind being tracked on SNS

Weak tracking The user does not mind if his name appears on the Friend list but he does not want his Friends to put a tag on their Profile linking to his profile.

No tracking The user does not want to be mentioned at all in his Friends' profile: no name, no tags, no photos.

V. IMPLEMENT PRIVACY FRAMEWORK

The proposed Privacy Framework is exhaustive and is able to cover all the possible case of privacy. However, normal users have to spend a lot of time, especially at the beginning, to understand and to configure their privacy settings. Fortunately, the level of privacy risks of each user can be determined by his usual behaviors and attitudes on SNS. There are five distinct prototypes of SNS users: Alpha Socializes, Attention Seekers, Followers, Faithfull's and Functional.

Alpha Socializes is people who use SNS to flirt, meet new people, and be entertained. They like to traverse Friend lists and put lots of comments on others' profiles and photos. As a result, their network and number of Friends are quite large but most of them are only Casual Friends. Alpha Socializes may also give to Friends their contact details such as MSN address or phone number so they can communicate easily outside the SNS. These actions can lead to disclosure of personal information and Security risks.

Attention Seekers are people who crave attention and comments from others. To get attention, they often post lots of photos, primarily photos of themselves and Friends in "suggestive poses, partying, drinking and portraying 'glamorous' lifestyles...". Their network is quite extensible; nonetheless they tend to have active online connection with only a few Friends. Due to the large number of photos, the Attention Seekers are the most susceptible to Reputation and Credibility risks.

Followers are people who join SNS to keep up with what their peers are doing. They often browse through Friends' album, occasionally exchange comments and update their profile. Compared with Alpha Socializes and Attention Seekers, users in this group are less likely to contact or meet people who they do not know. Consequently, most of their Friends are Best Friends and Normal Friends. There are many Followers on SNS; they have a moderate level of Reputation and Credibility risks as well as Profiling risks.

Faithful are people who typically use social networking sites to rekindle old friendships, often from school or university. They often leave their profile public so that old friends can find them on SNS. For them SNS are useful tools to strengthen existing offline networks rather than to create new, virtual ones. Due to the profile being public, the Faithful are easy victims of profiling risks.

Functional are a minority of people who tend to be single-minded in using SNS for a particular purpose, such as organizing parties, viewing photos, doing charity work. They are occasional users and generally log on for short visits. These users also suffer from privacy risks because they don't spend the time to learn about privacy settings and just leave their profile opened by default. By asking

ISSN (Online) 2278-1021
ISSN (Print) 2319-5940

IJARCCE

International Journal of Advanced Research in Computer and Communication Engineering

ICITCSA 2017

Pioneer College of Arts and Science, Coimbatore

Vol. 6, Special Issue 1, January 2017

users various simple questions such as “Why do you join SNS?” or “How often do you visit your page?” the framework would be able to classify them into appropriate prototype. Based on the characteristics of each prototype, we can propose to the user an appropriate privacy level that would give him enough freedom to do what he wants on SNS. For example, the main concern of a Faithful is to connect with old friends or distant relatives. Thus, he would be more comfortable with Soft Privacy, because with Healthy and Harmless data being public, it is easier for other users to find him on SNS.

VI.CONCLUSION

Social networking sites have become a potential target for attackers due to the availability of sensitive information, as well as its large user base. Therefore, privacy and security issues in online social networks are increasing. This survey paper addressed different privacy and security issues, as well as the techniques that attackers use to overcome social network security mechanisms, or to take advantage of some flaws in social networking site. Privacy issue is one of the main concerns, since many social network user are not careful about what they expose on their social network space. The second issue is identity theft; attackers make use of social networks account to steal victim’s identities. Social networking sites try to implement different security mechanisms to prevent such issues, and to protect their users, but attackers will always find new methods to break through those defenses. Therefore, social network users should be aware of all these threats, and be more careful when using them.

REFERENCES

[1] http://80legs.com/.
[2] Firegpg browser extension. http://getfiregpg.org/.
[3] CBCNews. Concordia bans Facebook access on campus computers 2008
[4] http://www.cbc.ca/consumer/story/2008/09/17/mtlconcordiafaceboo k0917.