International Journal of Advanced Computer Research (ISSN (print): 2249-7277 ISSN (online): 2277-7970)
Volume-3 Number-1 Issue-8 March-2013

Privacy and Security: Online Social Networking

Akriti Verma[1], Deepak Kshirsagar[2], Sana Khan[3]
Department of Computer Engineering and Information Technology, College of Engineering, Shivaji Nagar, Pune

Abstract

Online Social Networking (OSN) sites such as Facebook, Twitter, Google+ attract hundreds and millions of users. Such social networks have a centralized architecture wherein user's private data and user generated content are centrally owned by a single administrative domain that manages communication between its users. As a result, centralized social networks have gathered unprecedented amounts of data about the behaviors and personalities of individuals, raising major privacy and security concerns. This has put in demand for a decentralized social networking site that addresses the privacy and security issues.

Keywords

Online social networks, centralized service providers, privacy, security, identity manager, identity theft, data dissemination.

1. Introduction

The emergence of online social networks brought an era that changed the whole scenario of online information sharing. The earlier methods used included electronic mail messaging wherein the user could just pass on a text document or an image of a particular size. The size varied with the service provider. Some e-mail service providers allowed greater file sizes whereas some restricted to a small file size. When these social networking sites came into existence, users experienced new ways of sharing information as well as performing other online activities. These included online chat, video chat, sharing common interests, playing online games with friends, keeping touch with friends, near and far relatives, gaining information about them through their profiles and knowing their whereabouts. Therefore, by providing these exciting features at less cost these sites gained a lot of popularity. Out of the various social networking sites that exist today Facebook, Twitter and Google+ are the most popular [1] as they deal with millions of users worldwide. Although these websites have greatly increased internet usage, they have also created various privacy and security concerns in the research community. These concerns arise due to centralized architecture of these social networking sites. Most of the issues are due to the fact that these SNS (social networking site) vendors treat users as 'consumers'. This paper discusses two issues related to these sites, which are:

- Identity Theft
- Dissemination of user data by the service provider: Case study on Google+

2. Problem Statement

The present online social networks do not provide privacy and security to users. Their centralized architecture and techniques employed for online information sharing has left wide holes for online fraud, threatening users' lives. Therefore, to ensure and enhance privacy and security an architecture is needed that incorporates privacy principles and provides secure mechanisms for information sharing ,to its users.

3. Literature Review

Identity Theft

There have been various findings in the research community related to identity theft. In identity theft a person steals another person's identity and pretends to be that person by using his identity [2]. It has been found that millions of internet users are victims of this illegal online activity every year. It leads to great loss of time and money spent in identity repair and recovery. A research carried out in Australia, on Australian users of online social networking sites revealed that up to $3 billions are lost every year due to identity theft. A survey carried out in 2011 for identity thefts in social networking sites revealed the following statistics, as shown in Table 1. The table gives the percentage of identity thefts in each of the mentioned social networking sites [3]:

Table 1: Percentage Identity thefts

Social Networking Site	Percentage Identity Theft
Linkedin	10.1
Google+	7.0
Facebook	5.7
Twitter	6.3
My Space	2.7

The reasons behind this kind of a pattern are

- Users trust the social platform providers.
- The social platforms gather user information that is shared with the third party websites and can be viewed online by typing suitable queries on a search engine.
- Lack of awareness among users.

With the increase in use of social networking sites there has been an increase in online identity thefts. Inventions of these sites have created more and more opportunities for stealing identity and committing online fraud. When a user signs up for a social networking site such as Facebook, Google+, Twitter, even if he does not want to give his personal details to a stranger, his information can be viewed in the form of a profile by anyone who browses the internet. Therefore profile information of an individual OSN user can be easily obtained from search engines. When Facebook's privacy policy changed [4], the site's default settings were such that anyone who browses the internet could see status updates made by users, their photos and most of the profile information which includes details about a user such as real name, age, birth date, the language he speaks, his occupation and much more. In the same way anyone browsing for a Google+ profile can easily view it. Certain sites such as Flickr and Youtube share users' pictures and videos thereby providing a deep insight into a user's life including his friends and relatives. Some sites also reveal user interests and hobbies. With this type of information easily available on internet a user can be identified and a fake profile of the user can be created. The following scenario illustrates how someone trying to impersonate another person can become successful.
If Bob wants to impersonate Steve, Bob finds the following information about Steve while browsing:

- date of birth
- language
- place of employment
- address
- place of birth
- phone number
- mother's name
- hobbies, interests
- picture of Steve
- Steve's contacts

He can create a fake profile of Steve and can send invites to all the people in his friend circle. People, identifying it as Steve's profile, accept the invitation. After adding Steve's friends, Bob can chat with his friend's and read their posts and messages through which he can gather more information about Steve as well as his friends. During chats, Bob can stealthily steer the conversation in a direction that enables him to find out everything he wants to find about Steve. After gaining this information, Bob can completely impersonate Steve on web and can invade into his personal and professional life. Since Bob has access to Steve's friends' profiles, he can put viruses that replicate through messages on his friends' systems. These viruses can record login details while they sign up into their social networking site accounts.
These online practices can lead to the following consequences:

- A person can log into someone else's mail or social networking site account without their permission and access or change password protected information on purpose.
- Menace, harass or offend someone, for example posting threatening messages on someone's Facebook wall posts.
- Trick someone to get something out, for example using someone else's credit card number to buy stuff without their permission.

Dissemination of user information by the service provider

The centralized architecture of the current online social networks is such that, all user data is stored in a central server. Therefore all the data is in control of a single administrator who has the capability to use any user's information to serve his interests such as earning revenue by selling this data to other firms. All social networks including the popular ones such as Google+, Facebook, Twitter have a centralized design. In this paper, dissemination of user information by Google+ has been discussed, stating the type of information it collects.

Dissemination of information by Google:
Google uses its various products such as Google Adsense, Google Analytics, Picasa web albums, Google Maps and many more to collect user information. Google+ is a social platform provided by Google to which it has integrated its various services and has transformed its Google search engine into this platform. Therefore Google now collects enough information on each of its user, such that it can even create individual dossiers of its users. This information that Google collects can be classified as:

- Personal information: It includes user's real name, age, birth date, birth place, etc. This type of information can be collected from various social networking sites developed by Google such as Google +, Orkut.
- Sensitive information: This includes user's medical information, his language, gender, political and religious beliefs. This information can be combined with personal information.
- Non-personal information: This information is collected from user's online activities such as net surfing using the Google search engine, visiting websites that are tracked by Google or the websites owned by Google.

This information gives an idea about user's interests and preferences [5].
Google states in its privacy policy that it protects this information and is in compliance with the safe harbor program, but gives no guarantee whether a change in its privacy policy will provide the same kind of protection to this information or whether user information will be protected. It has been predicted about Google that once it acquires sufficiently large information on most of the population of the world it can change its privacy policy and can share users' data to expand its business. It has been stated in its privacy policy that in case of an acquisition or a merger with another firm, user information can be released. Although this act of Google will be made noticeable to its users, this may lead to severe consequences for a user, such as identity theft and other cybercrimes, threatening his life [6]. This provides us with the major drawbacks of using the principles of a centralized architecture.

4. Proposed Solution

The privacy issues related to the current centralized online social networking sites inspired researchers to direct their research towards designing mechanisms for preserving privacy in online social networks. This led to the emergence of decentralized social networks. In a decentralized social network, user's data is not with a single entity [7]. A typical decentralized network consists of distributed servers that are either owned and administered by the user governing his own data or are owned by an user trusted organization. Therefore decentralized social networks are implemented on a distributed management platform, where communication follows a peer-to-peer (P2P) approach [8]. Based on these principles of decentralization many social networks came into existence out of which Diaspora became the most popular [9]. Diaspora's advantage over other social networks is its privacy preserving design that incorporated all the principles of a decentralized architecture [10]. Unlike the existing centralized social networks, Diaspora is a federated social network that consists of a number of servers, known as 'pods' or 'seeds' that are distributed over the network. These 'pods' contains all the data related to a user who is either the administrator of the pod or just a user of the pod that is administered by a trusted party. Therefore, Diaspora is a distributed decentralized social network that protects user's privacy and provides secure communication. Diaspora uses the following protocols:

1) Salmon Protocol
2) Ostatus
3) Webfinger
4) Activity Streams
5) PubSubHubBub
6) Advanced Message Queuing Protocol
7) HTTP Secure

Proposed Architecture
In real life, verbal communication between two people happens in such a way that the messages that are being exchanged are only received by the communicating parties. The third party (everyone except the communicating parties) are unaware of the message contents. Data is disclosed to the third party only if the communicating parties agree to release the message contents. This is indicated by the figure shown below.

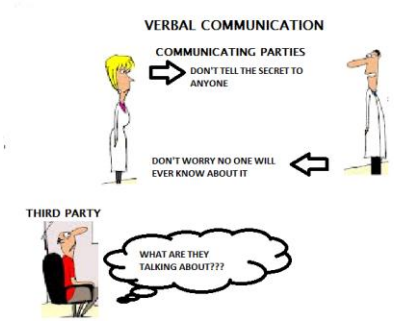

Figure 1: Verbal communication

This idea behind communication is very important in order to preserve a person's privacy and security during information sharing. To incorporate the principles behind real life verbal communication into online data sharing, the concept of a "Freedom Box" is used. A "Freedom Box" is a device that runs an operating system and can act as personal server [11]. This "Freedom Box" can be used to implement distributed and decentralized social networking using Diaspora as the social networking site. In this architecture the "Freedom Box" is configured to handle the Diaspora database such that all the user related information is stored on the database. This information includes user profile information, information regarding posts and messages sent to other Freedom Boxes or received from other Freedom Boxes. Therefore, each "Freedom Box" is a pod owned by a single user. In our architecture we are not including the concept of a 'community pod' because here we regard "Freedom Box" as a device that is small, handy like a cellphone, and easily available to everyone who wishes to have an independent server. This kind of a design enhances privacy and security. This architecture is shown below

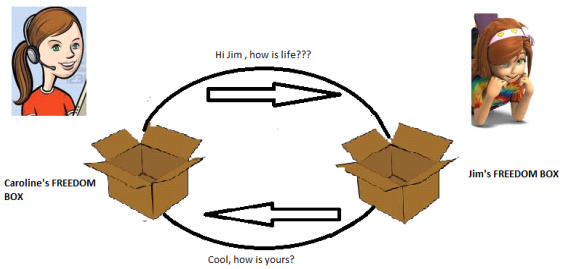

Figure 2: Decentralized social networking using ""Freedom Box".

Here communication is taking place between two Freedom Boxes that are running Diaspora server. The phenomenon of information sharing in such architecture is illustrated below.

Let us consider Caroline is hosting a server at box1 and Jim is hosting a server at box2.

- In order to start communication, it is important that the two Freedom Boxes connect to each other. To connect her server with Jim's server, Caroline will have to locate Jim's profile. If Jims profile information is already stored on Caroline's server she can just use the search feature on her home server to locate Jim's profile. If this is not the case Caroline will have to know Jim's Diaspora handle to access her profile information. This handle is a user identifier and can be shared through any other means of communication such as telephone, short message service or e-mail, according to the convenience of the communicating parties (in this case Caroline and Jim). This Diaspora handle looks like an e-mail address such as caroline@dias.org jim@diaspora.org [9].
- Once the connection is established using the Diaspora handle, Jim's profile information is copied to Caroline's server. This happens because whenever a Diaspora handle is used to connect to another server, the source server queries the destination server and fetches the profile information from that server.
- Now, communication starts between Caroline and Jim via messages. Whenever Caroline posts anything on her profile's user interface, it is pushed upstream by the protocol PubSubHubBub.
- The activity streams protocol attaches the identifying information to the message and to all its subsequent updates. This information includes the name of the publisher, the time when it was written and the kind of message (such as an article, a post or a status).
- The queuing and routing of the posts and messages is handled by the advanced message queuing protocol [12].
- The HTTPS protocol uses cryptographic techniques such as RSA (random sequence algorithm) and digital signature for encoding the shared data and authenticating the sender. Therefore, the data shared between

Caroline and Jim is encrypted and Jim is assured that the information he receives comes from Caroline [13].

- Using the information attached by the activity streams protocol, the Diaspora software running on Jim's server displays the published data.
- Jim, on viewing Caroline's post wishes to reply to her. The reply sent by Jim is put upstream by the salmon protocol and undergoes the same procedure [14].

5. Comparison

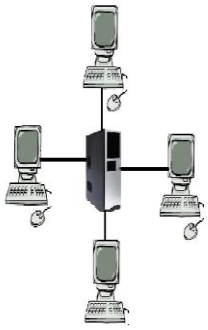

Figure 3: Centralized architecture of current social networking. Here the computers imply 'nodes' and the 'CPU' implies the central server which contains all users' data

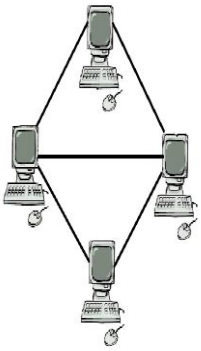

Figure 4: Decentralized and distributed architecture where the computers imply 'pods'

- In centralized architecture, Fig. 3, the service providers have full control over users' data that includes personal, non-personal, sensitive information whereas in the proposed decentralized architecture, Fig. 4, a user is given full ownership of his data.
- The lack of user authentication mechanisms in the current centralized social networks make them prone to the following attacks:
  Man-in- the-Middle attack: In this attack the public key shared by the communicating parties is intercepted by a third party, which causes the communication to take place in such a way that the communicating parties get an illusion that the communication is happening between them, whereas the communication actually happens between the communicating parties and the third party [15].

  Evil twin Attack: In this attack a person impersonates another person on web to satisfy his evil intentions such as to threaten someone or to gain access to sensitive information or resources which he is not authorized to access otherwise. This attack is similar to identity theft in nature and can impose server consequences on users [16].
  To prevent these kinds of attacks in the proposed architecture, Diaspora uses HTTPS protocol which provides sender authentication through digital signatures. Since two users cannot have the same digital signature it becomes extremely difficult to impersonate another person or to communicate on behalf of someone else.

- A commonly found, illegal online practice in the present social networks is cyber bullying. In cyber bullying a person claims to be a different personality by creating a fake account, providing incorrect details about himself. This cyber crime is different from identity theft, as here a person may or may not impersonate another person. He might just claim to be of a different nature, sex, age and change other details that will help him disguise himself. Such user may then connect to other people on internet, which may prove to be harmful for them [17].
  In the proposed architecture, for a connection to take place between two servers it is important for the owners of the servers to share their Diaspora handles with each other. The various means by which they can do this are e-mail, telephone, short message service, video chat. Since sharing happens in this way, the owner gains sufficient information on the other person hosting another server, to decide whether the connection should take place between them, thereby avoiding cyberbullying.

- The service providers of the current online social networks use social graphs to represent user information. If these graphs are released to other companies or if a third party is able to gain access to them by methods such as hacking, all the information

about a user such as his relations and his internet behaviour is revealed to the third party.

In the suggested architecture no social graphs are maintained as the nodes are not tied up. This means every time a new connection is established between two nodes using Diaspora handles.

6. Conclusion

The Architecture of the present centralized social networking sites is such that they do not ensure privacy and security to users. Therefore, there is need of an architecture which is privacy and security preserving such that a user feels like a 'user' and not a 'consumer'. These requirements of a user are catered to by a decentralized distributed architecture. Since the proposed architecture is decentralized and distributed (uses "Freedom Box" as servers) and uses Diaspora as the social platform, which enhances privacy and security through its various protocols, this architecture preserves privacy and security of a user and must be employed by people for information sharing.

Future Scope

Various devices such as cheap plug computers have been launched in the market. Cheap plug computers include Dreamplug [18], D2plug, Raspberry pi [19]. These devices can be used as a "Freedom Box" .These devices consume less power and provide large uptime. They are cheap and easily affordable by a common man. College Students with Distributed systems and Networking as their project domains can create Diaspora applications that will work with this architecture, using ruby on rails as the framework.

References

[1] Wikipedia,Online at http://en.wikipedia.org/wiki/google.
[2] Wikipedia, Online at http://en.wikipedia.org/wiki/identity_theft.
[3] Javelin research strategy, Online at http://www.javelinstrategy.com.javelin .
[4] Rafael Sofaer, Maxwell Salzberg, Ilya Zhitormirskiy and Daniel Grippi, "The Anti-Facebook," IEEE Spectrum, pp. 47-51, 2011. Available: http:// www.spectrum.ieee.org.
[5] Stefanie Alki Delichatsios and Temitope Fonuyi, "Get to Know Google Because They Know You", Ethics and law on the electronics frontier, pp. 2-20, 2012.
[6] Google Privacy Policy, Online at http://www.google.co.in/policies/privacy/, (as of 14 December, 2005).
[7] Shirin Nilizadeh "Cachet: A Decentralized Architecture For Privacy Preserving Social Networking With Caching", Proceedings of the 8th ACM International Conference on Emerging Networking Experiments and Technologies Co NEXT, December 10-13.2012, Nice, France.
[8] Oleksandr Bodriagov "Encryption for peer-to-peer Social Networks", proceedings of IEEE international conference on social computing, pp. 1302-1308, 2011.
[9] Ames Bielenberg "The Growth Of Diaspora –A Decentralized Social network in the Wild", IEEE, pp.13-18, 2012.
[10] Wikipedia, Online at http://en.wikipedia.org/wiki/Diaspora.
[11] Eben Moglen, "Testimony of Eben Moglen", U.S. House of Representatives Committee on Energy and Commerce, Software Freedom Law Centre, December 2. 2010. Available : moglen@softwarefreedom.org.
[12] Wikipedia, Online at http://en.wikipedia.org/wiki/Advanced_Message_Queuing_Protocol.
[13] Wikipedi, Online at http://en.wikipedia.org/wiki/HTTP_Secure.
[14] Wikipedia, Online at http://www.salmon-protocol.org.
[15] Wikipedia, Online at http://en.wikipedia.org/wiki/Man-in-the-middle_attack.
[16] Carl Timm, "Seven Deadliest Social Network Attacks", series X, Syngress, Elsevier Inc., pp. 62-82, 2010.
[17] Carl Timm, "Seven Deadliest Social Network Attacks", series X, Syngress, Elsevier Inc. pp. 99-111, 2010.
[18] Wikipedia, Online at http://en.wikipedia.org/wiki/DreamPlug.
[19] Wikipedia, Online at http://en.wikipedia.org/wiki/Raspberry_Pi.

Akriti Verma, born in Pune, India, on September 10, 1991, is a Final year Engineering undergraduate student of Information Technology at the College of Engineering, Pune, India.