EAI Endorsed Transactions
On Industrial Networks and Intelligent System
Research Article ICST.ORG

Security and Privacy in Online Social Networks: A Survey

Sudarshan Kudlur Satyanarayana[1*], Keshav Sood[2], Yuan Tao[3] and Shui Yu[4]

[1]Deakin University, Australia, skudlur@deakin.edu.au
[2]Deakin University, Australia, ksood@deakin.edu.au
[3]Anhui Earthquake Bureau, Hefei. School of Computer and Information Hefei University, China, taoy89@126.com
[4]Deakin University, Australia, shui.yu@deakin.edu.au

Abstract

The Online Social Networks (OSN) open a new vista serving millions of users and have reshaped the way people interacts. Unfortunately these networks are an emerging platform for cybercrimes such as sending malicious URLs, spams, etc. which causes a huge financial loss and social damage. In this paper, it is reviewed that OSN is a new cybercrime platform for threating agents because of the security pit falls in the existing centralised architecture and their driven functionalities. This article surveys the detailed analysis of the current state of the Online Social Networks in perspective of security and privacy issues. Additionally in this paper we presents various types of attacks that can be mounted via OSN and the defence measures against each of the attack. This literature also highlights on the types of vulnerabilities and threats in OSN. Significant threat categories and risks associated as well as a scope to circumventing these threats and vulnerabilities are also introduced.

Keywords: Architecture of OSN, Online Social Networks, OSN Attacks, OSN Vulnerability and risks.
Received on 30 October 2014, Accepted on 30 October 2014, published on 09 December 2014

doi: 10.4108/ inis.1.1.e3

1. Introduction

Online Social Networks (OSN) are based on certifiable social connections [1] and give their clients wide options for the virtual-communication such as Facebook, WhatsApp, Viber, Vchat, Skype, Google Plus, Twitter etc., help people in the society to share news, reconnect with each other, share ideas and interests, build new relationships, conduct conferences and also reconnecting with old friends even though they are located in different geospatial regions.

In recent years the Online Social Networks have been so popular that about 300 or more OSN exits in enterprise domain serving more than billion [1]. This in turn means that the OSN database stores a huge amount of private and sensitive user data information that is intended to a specific audience/s. Adversaries may intend to use the data in any false way or they may act as a threat agents [1]. In other words, OSN is an emerging platform for adversaries for various of reasons such as cyber-attacks, stalking, reputation slander, personalized spamming, and phishing.

Popularity and usage of OSN equally give rise to critical issues of cybercrime and threats pose a new door for vulnerabilities. It is reported in April 2013 that hackers broke the twitter accounts and spread rumours about president Obama. London riots in 2011 is another example of threats caused via OSN [2]. Researchers revealed that OSN has brought a simple but new way for communities to connect each other whereas it also brought a new domain of innovations and challenges to avoid the threats and concerns being caused by OSN. Various perspective are being focused to mitigate the critical dangerous of OSN, including; understanding user behaviour, investigate traffic activity, popular user investigation, detecting source nodes of rumour and structure characteristics of social network, etc. [3]. This linear rise in OSN and cybercrime have brought critical concerns among researchers in industry, education

*Corresponding author. Email: skudlur@deakin.edu.au

community, and government urge to deeply analyse the pit falls in the existing OSN structure.

In this research effort Section 2 highlights the current architecture and main functionalities of a typical OSN. The privacy and security concerns are discussed in section 3. A thorough perspective of types of adversaries, threats and vulnerabilities are presented in section 4 and 5 respectively. Section 6 introduce breaches in OSN. Section 7 discuss the attacks and current defensive measure in OSN. Finally section 8 draws some conclusions.

2. Architecture and Functionalities of OSN

The fundamental aspect of any architecture (particularly OSN architecture) should be able to maintain the confidentiality and privacy of user information and user related data. Additionally the key point of concern while designing any architecture is to ensure the designs ability to be scalability, security and reliability. With our extensive review of OSN existing architecture we determined that the OSN architecture is classified into two broad categories. Client- Server Architecture and b) Peer-to-Peer Architecture. [4]

2.1. Client-Server and Peer-to-Peer Architecture

All of the present days OSN is based on the centralised Client-Server architecture or in other words Web-Server architecture. All the user data and the interactions are flowing through the server and vice versa. The functionalities such as storage, maintenance and access control are being controlled by the centralised server of OSN service providers like Twitter, Orkut, and Facebook and other service providers. Figure 1 illustrates a typical architectural view of Client-Server architecture [4]. As the control is centralised the network management is easy however it brings another critical issues primarily where the security is concern. The bandwidth and computation is the another bottleneck in Client-Server architecture.

An immense research has being carried out to convert the present client-server architecture o into the peer-to-peer architecture shown in figure 2. In this decentralised architecture, which relies cooperation of each of the users connected in and with the OSN, user's personal spaces are put away and kept up distributive [4].

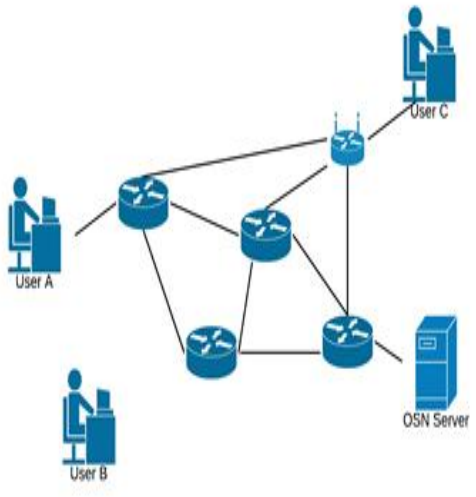

Figure 1. Client- Server Architecture

By supporting the immediate trade of data between applications, between users who have met before or between contiguous nodes of a mesh network, a P2P can exploit real social networks and geographic closeness to help nearby administrations when Internet access is occupied. Maintain privacy and autonomy from another service providers again is a big issue. Guaranteeing availability of data even when the data owner is not online is also a problem in this decentralised peer to peer architecture [5].

For the P2P architecture, offering a few functionalities (e.g., global search) of OSNs in an appropriated way is a challenging issue [5].

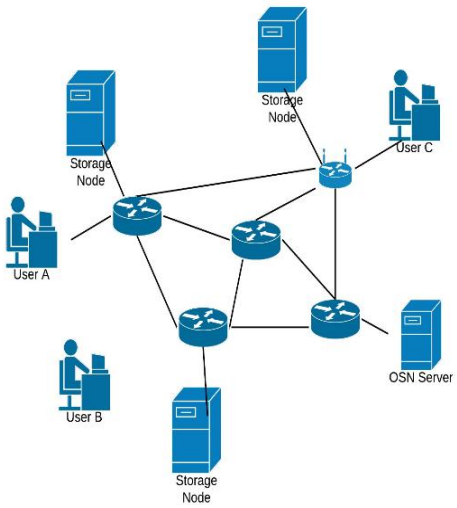

Figure 2. Peer-to-Peer Architecture

2.2. Functionalities of OSN

The objective obtain in-depth view of functionalities is to understand and to provide security features and flaws in the existing architecture and their driven functionalities. Also to become fully aware with the functionalities it is necessary to understand the flow of communication between end users.

User interacts within OSN in number of ways. Firstly users can leave the public message in the form of writing text, uploading video and audio file or even as the photos using there allocated personal space. Another is users can leaving private messages. Meanwhile the clients for real time application such as online games can also use OSN.

In the existing OSN, interactions are not restricted between the existing relationships but also enable to form new relationships. In order to find an unknown contacts the OSN uses two approaches: Global Keyword Search and Social Graph Traversal [6]. Global Keyword Search is used to find the unknown users by performing global keyword search. For example a user can find an unknown person in Facebook by doing search in the search bar. On success the searching entity provides the list of the similar users/search. Customize search using the Global Keyword Search is the search option in which search is based on various search components like name, country, location etc. In Social Graph Traversal Search OSN user can search the person by traversing through the friend's list of the other OSN users with whom they are in relationship with. This search is nothing but traversing through the friend list of the other user within a single OSN. Users can specify the policy that suggests that whom are allowed to traverse through his/her friends list.

In summary OSN functionalities or characteristics can be understood in two ways i.e. the ways OSN communicate or provide information; a) to provide the digital representation of the users and their relationships b) provides the networking services for sharing the information and, c) enables the messaging for the users [7]. Additionally, important concepts most widely used in OSN are: Digital Personal Space and Digital Social Space. Digital Personal Space is the collection of the individual users profile and connected links. The collection of the Personal spaces is called Digital Social Space [1].

3. Privacy and Security in OSN

The following sub-sections highlights all the necessary security features such as Integrity, Confidentiality and Availability in OSN.

3.1. Data Integrity and Authenticity

From network side OSN is mainly a representation of a real life Social Network or off line Social Network [1]. The data stored in the OSN can be modelled into an online social graph; there exist one-to-one mapping from real life social network to the online social graph [3]. Data integrity in OSN refers to the consistency that should be maintained, that is any deviation from the real life social network should be detected as an attack and preventive measures should be taken to evade this attack.

When speaking about Integrity two type of attack prevails: a) forging Nodes/identities and b) forging social links/connections [1, 8]. Forging node is the one where adversary can forge the node or perform identity theft. For example, an adversary can create fake profiles of the famous people or the brand to defamation people from their reputation. In another adversary type that exists, the attacker can create multiple fake identities of intended victims. Therefore the user should be assured of the communication authenticity and must be confident that the communication is legitimate.

3.2. Privacy and Confidentiality

Privacy and confidentiality refers to the user private and personal data that should not be disclosed illegally and to ensure that is to be used by authentic user/s only. The privacy is divided into a broad categories i.e. User's Identity Anonymity, User's Personal space Privacy and User Communication Privacy.

In the first category, since many OSN exists across the world the protection of the user identity also changes from different OSN. Some OSN like Facebook uses the user's real name to represent the user to the rest of the OSN. Some dating sites uses the user's first name for representation. Some OSN exists where a random identifier is used for the representation whereas name and the contact details are not used.

In the second category, User's Personal space Privacy, the user profile visibility also varies form one OSN to another. OSN like Friendster makes the user profile visible to everyone regardless of whether the person searching has an account or not. Some OSN like MySpace and Facebook allows the user to specify to whom their profile should be visible, that is whether their profile is visible to public or friends only. Most of the OSN has feature that the friend list is visible to people who can view there profile but there are some exceptions like LinkedIn where the user can hide the friends list.

Lastly in User Communication Privacy, along with the user personal data is disclosed on the digital space of the user. The user can also disclose their personal information to the network providers and also to the OSN provider. These data may include IP addresses, message that were communicated, other profiles visited and so on. An extra measure should be taken in order to maintain the privacy of the communication.

In a nutshell, client's security necessity is twofold. To begin with, unapproved elements (i.e., who are not conceded access to the private information) should not take in the substance of the private information, which uncovers recognizing data of the information holder

(client). This part of information protection intimates information confidentiality what's more the holder's obscurity, and specifically prompts the requirement for access control. The client might just allow access to data on a client straightforwardly, and the right to gain access control must be as fine-grained as every private data thing must be independently reasonable. Second, unapproved elements should not have the capacity to connect numerous private information to profile the users, demonstrating that the put away or transmitted private information ought to seem arbitrary and release no helpful data. This angle is basically the unlink ability necessity.

4. Types of Adversaries

In the OSN communication, many parties/groups or individual for different reasons attempt to steal the user information. Typically such entities are called attackers. In the following literature we mentioned the types of attackers or adversaries.

a) Insider Attacker

These types of attackers appear to be legitimate users but perform malicious activities staying inside the network. Some examples are malicious user in the OSN and malicious user who has the network activity to perform malicious activities like eves dropping or can perform man in the middle attack. The main threat to the OSN is the unintentional *Insider Attack*. In this context the adversaries try to manipulate the user of the OSN to perform the malicious activity without the user being aware of the attack being carried out. This type of threat uses the human psychology and cognitive bias of the user [9]. Adversaries carry out this attacks in two ways or stages; Single Stage Attack and Multi Stage Attack.

In the first Single Stage of Attack the intruder obtains the sensitive information from OSN as a result of the exploit of confidential information and uses such information to generate an attack. But the adversary do not use the result of such exploit to carry out further or subsequent exploit but directly use to attack the user or the OSN. While in Multi Stage Attack, the attacker use the information obtained from the exploit to further carry out various exploits. The result of the exploit can be for the matter of minutes or can last for weeks or months depending on how the attacker uses the result of the exploit [10].

b) Outside or External Attackers

The attackers in this case are not the legitimate users of OSN, however, then can perform attacks on social networks or on network architecture. The result of such an attack or the attacks lasts for may be a matter of minutes or for week or months depending on the way attacker uses the information to carry out the attack [10].

5. Threats and Vulnerabilities of Online Social Networks

In the following section we have discussed the vulnerabilities and risk associated with such threats and attacks. Table 1 also highlights Threats and Vulnerabilities of Online Social Networks.

5.1. Privacy Threats

a) Digital Dossier

The advancement in the data mining results in cost required for the data storage is reduced significantly, it is possible for a third party to create digital dossier of all the personal information of the OSN user that is displayed on their profiles. The major vulnerability that exists in the OSN is that all the personal and sensitive information is available through the profile search [1, 12]. Revealed information on the OSN can be exploited by attacker to blackmail or embarrass or even damage the identity of the profile user. For example an intruder can corrupt a user profile as in the present day the employer's tries to view public profile of the selected candidates.

b) Face recognition

OSN uses face recognition [13] to identify people in the OSN by uploading the images to their profiles. These uploaded images can be used to create fake profiles across different OSN without the user being aware of the fact that his/her image has been used to create the profile. The attacker can use the stolen images to link the users across different OSN.

c) Content Based Image Retrieval

Almost all OSNs that are present today have not implemented any security measures against the Content-Based Image Retrieval. CBIR [14] is an advanced technique that can be used to match features such as identifying some aspect of the room from a large database of images and thus obtaining the location of the users. This leads to disclosure of the location of the OSN users leading to stalking, blackmailing and unwanted marketing and all others attacks that might come across from the location disclosure.

5.2. Traditional Network and Information Security Threats

As the social networks are growing widely, it encouraged attackers to create unsolicited messages called Social Network Spams [15] to produce overload in the traffic. These spams produces traffic overload and can introduce

Table 1. Threats and Vulnerabilities of Online Social Networks

Types of Threats	Vulnerability	Risk associated
Privacy threats Example: *Digital Dossier, face recognition, Content Based Image Retrieval.*	Personal and sensitive information is available through the profile search, create fake profiles.	Blackmail or embarrass or even damage the identity of the user profile by attacker, unwanted marketing
Traditional network and information security threat Example: *Spamming, Cross site scripting, worms and viruses*, Denial of Service Attack (DOS	Create unsolicited messages, traffic overload	Loss of trust or difficulty in using the applications, introduces phishing and unwanted site traversal
Threats relating to identity Example: *Phishing*	Exploit sensitive information	Retrieve personal sensitive information, credit card details and thus which might cause the financial loss or even personal losses such as reputation.
Information retrieval threats	Retrieve the sensitive information pertaining to the restricted group	Information leakage, phishing attack or the spam attack.
Profile hijack by identity threat	Create fake profile to impersonate the person or a known brand.	Creating bad reputation, financial loss

loss of trust or difficulty in using the applications, and also introduces phishing and unwanted site traversal. Additionally the OSN also has many widgets provided by the third parties and these widgets are not well verified [15]. The weak verified widgets introduce the vulnerability for the cross-site scripting and also can introduce worms and viruses when the OSN user clicks on it. The attacker can use these vulnerabilities to perform Phishing attacks, compromise the account of the user and also send the directed spams to the user accounts. The most important attack that can be introduced by this vulnerability is Denial of Service Attack (DOS).

5.3. Threats Relating to Identity

Due to the availability of the sensitive information in OSN the phisher can easily exploit these information to carry out phishing attacks at a very high success rate. OSN is also vulnerable to the scripting attacks that help the adversary to embed the phishing links. The risk that are associated of these phishing vulnerability is that it allows the attacker to retrieve the personal sensitive information, credit card details etc. which might cause the financial loss or even personal losses such as reputation [16].

5.4 Informational Retrieval Threats

The Social networks privacy can be exploited when attacker becomes friends with the user of a restricted group. Once he/she has become friends with the person he can easily retrieve the sensitive information pertaining to the restricted group.

Attacks that can be conducted due to this vulnerability is information leakage, phishing attack or the spam attack [17].

5.5. Profile Hijack by Identity Theft

The attacker can create fake profile to impersonate the person or a known brand. If he knows the personal details of the user and create profile to mimic the users. The main risk associated with this vulnerability is the loss of reputation of the person or brand under attack which eventually may causes financial loss.

6. Breaches in OSN

Participating entities in OSN are typically classified as: Users, Service provider and Third party applications. In this section we briefly detail the breaches that might generated from the participating entities in OSN to disrupt the privacy of the user data. As shown in figure 3, the pie chart, depicts that 90.8% of the OSN users uploaded their personal images, 50.8% disclosed their date of birth, and 39.9% users disclosed their phone numbers and 87.8% users uploaded their current address. This abundance information make the attacker easy to exploit and obtain personal information and carry out the attacks. Further we have briefly described the breaches by different entities in flowing sub-sections.

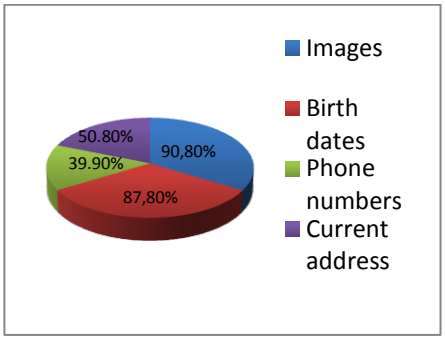

Figure 3. User disclosed information in OSN

6.1. Other User Breaches

The OSN provides communication between friends. To facilitate this functionality OSN service providers restrict unconfirmed access. The OSN provides the functionality that all users in a particular user friend list can view the images and other information pertaining to the owner of the data and also blocking the access to all others.

The OSN friends list is nothing but a graph where when two users becomes friends their nodes are connected to each other's [3]. It is merely a link that exists when two user have agreed to establish in a particular OSN.

This particular link is vulnerable to information stealing attacks. Based on this vulnerability there can be two attacks that can be mounted: Same site profile cloning [6] and Cross site profile cloning [6].

a) Same-Site Cloning

In Same-Site profile cloning the adversary can steal all the personal information from one legitimate user and then duplicate the profile within the same OSN. Further on using the same duplicate profile, attacker send the friends request to the victim's friend. The person when receiving the request can see the request from a known person and accepts the request providing all his personal sensitive data to the attacker.

b) Cross-Site Cloning Attack

In Cross-Site cloning attack the attacker performs the identity theft and steals all the personal information from one user on OSN-A, and then creates a profile from the stolen information on OSN-B. Once attacker creates the profile on different OSN he/she can send friends request to all the friends in the victim's friends on OSN-A to join OSN-B. Users on receiving the request thinks the request as legitimate and then joins the other OSN shared all their information.

There is no defensive measure that exists today against this attack and is it is the user's responsibility on which user they want to add as friends. However, by using the *CAPTCHA* we can reduce the number of cloning attacks that might take place by automated scripts.

6.2. OSN Service Provider and Breaches

Existing OSN are built on the centralised client server architecture. This implies user of the OSN to trust OSN service provider on storing and protecting their personal information. Since the service provider has the access to the personal information of the user the service provider can certainly benefit from this by monitoring and sharing this personal information for example, service providers can use the user data or the photo uploaded for an advertisement. This aspect has raised many questions by the researchers on usage of the data by service provider.

The researcher in present days has proposed many solutions for this vulnerability. The user should be given the fine-grained access to control who can access his or her information. To specify user policies the OSN can store the information by encryption so that even the service provider cannot access or view the user information unless the user explicitly specifies rights in his policies. The other solution is to make use of a decentralised architecture so that the data is not stored in centralised location [18]. Other solution provided by the

researchers is to use smart clients and untrusted. The server stores the encrypted data and this data is available only to the user has given rights in his policies [18].

6.3. Third Party Application Breaches

Extensive growth in OSN networks and users, the user demands for more functionalities to add to the OSN is continuously increasing. But these functionalities are added by the Third Party and are always untrusted even though they reside inside the OSN. These Third Party applications require the users to provide the access to their personal data to provide their functionalities. For example, a horoscope application requires the user's date of birth to provide the horoscope details.

But the main vulnerability exists that user or the service provider do not knows which part of the user personal information is required by the applications. As a result the user rely a trust on the application in a manner that the application will use users information in a sophisticated manner and do not disclose their personal information. Adding to the vulnerability there is no mechanism that is in place in the current OSN to monitor how the third party application are using the data. Thus making it very easy for information misuse by the third party application.

One of the defensive measure specified is the use of XBook [19]. The XBook uses the flow-based model, which controls the data usage by the untrusted third application. Any communication that occurs between the user and the third party application must occur through XBook Whenever the user wishes to add the application the XBook provides the list of information that the application can use. Further the XBook monitor whether the application is using and publishing only the part of information that the user has agreed or that the application can share. This provides the network monitoring but almost none has provided that interface or communication establishment between applications to XBOOK. Hence this is also not a trustworthy mechanism that exists to solve the third party breach issue.

One solution that the OSN Service provider can provide is to increase the privacy policies so that the applications should have to get the approval from the user before accessing the information that is hidden and only visible to user or his friends.

7. OSN Attacks and Defensive Measures

In the following section, we discuss several kinds of attacks in OSN. Possible methodologies of mounting an attacks and respective defence measures are classified and discussed below.

- *De-Anonymization and Spam/ Active and Passive*
- *Malware*
- *Reverse Social Engineering*
- *Distributed Denial of Service*
- *Sybil*

7.1. De-Anonymization and Spam

a) De-Anonymization

The attackers used de-Anonymized attacks to breach the anonymized data and thus attempts privacy breaches. Scientists and researchers are paying much attention to reveal the anonymised structure of social network assumed to be a new attacker in the environment. Such anonymised attackers mount attacks in various ways. Report released by Symantec reflects that about 50% of the cyber-attacks mounts through the Viruses, Trojan horses and worms, and 18% of the attackers are targeting OSN in particular for their attacks [20] as shown in figure 4. Based on the way attack generate, they are classified in two distinct categories; Active Attacks and Passive Attacks.

In active attacks the attackers registers and creates few accounts thus creating the link patterns and then connects this link to the targeted users/victims. Once this is done using de-Anonymization the attackers can re-identify the targeted users [21].

In passive attacks the attackers do not create any accounts or nodes for de-Anonymization but instead exchanges the structural information among the small coalition of groups or friends and then identifies the sub graph, which makes the friends to identify themselves [22].

An intruder picks a self-assertive set of OSN users whose privacy it intends to breach, makes a little number of new user accounts with edges to these targeted victim's and makes a pattern links among the new records with the objective of making it (intruder) emerge in the anonymized graph structure. This attack involves creation of new nodes whose edges may help to re-identify the existing nodes.

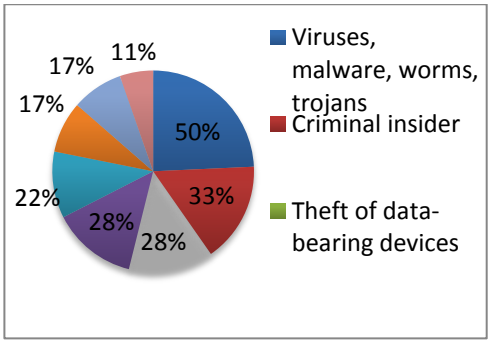

Figure 4. Cyber-Attack trends

As de-Anonymization attacks requires large data to carry out the attacks, the simple defence against this attack can be obtained by making the acquisition of data however is

more difficult. Since the present day OSN is centralised, the major protection are being provided by the services providers on the server side than on the client side protection. In the following literature we are highlighting about spams in online social networks. Spams can happens in two ways Context Aware spamming and Broadcast spamming.

1. Context aware spamming

Spammers takes advantage of the common context (like location, interest, workplace etc.) between users on OSN to obtain high click through rate [22] to enhance their ability to select the set of target and send spams. Context aware spams commonly send to victims via *SMS*, emails, advertisements in OSN, malicious web link or pages. These spams may be more vulnerable as it may install malicious software or could steal sensitive information like passwords. The most likely medium of spam propagation is email, as per the recent study about 70% of the OSN users display their email addresses in their profile making it easy for carrying out the attack. Context aware spams are of three types [22];

- Relationship based attack; this attack uses the friend-to-friend relationship and no other attributes will be used.
- Unshared attribute attack: this attack uses the friend-to-friend relationship along with the attribute which unique for one person like date of birth.
- Shared attribute attack: this attack uses friend-to-friend relationship along with the common visible attributes such as hometown.

2. Broadcast spamming

This type of spam is an untargeted spam but interrupts the public communication to obtain the information [2]. For example, consider in Facebook the spammer can pollute the image tagging such that when a person clicks the links he/she can be redirected to the spammer's assigned link. Another example is of phishing. In this user/s are getting the ads on Facebook, the ads are posted by performing the context aware spamming and once the user clicks on add this leads user/intended victim to a malicious website.

Since the information is very easily available in the OSN the phishing carried out on the OSN is four times more effective than the traditional phishing. With the information theft exploited by the attacker he can impersonate as the victim and send the spoofed messages to his friends/another users/groups with the link. When the users click in the link the malicious software immediately affects them.

7.1.1. Defensive Measures against Spams and Phishing

In this section we survey existing mechanism that are developed to prevent spam and phishing in OSN.

a) Identification or detection

The identification-based methods carry out in two phases, in the first phase the user can manually identify the spams or the user can perform pattern-based classification. In the second phase the system will validate the classified spams in the first phase and then delete the spams later result is computed and displayed to the user.

b) Demotion based method

In demotion based method design [3] the framework to lessen the ranking of content prone to be spam. Rank-based techniques attempt to give better results ordering for interfaces that create arrangements of results. Preferably, these orderings are both more faultless and more impervious to spam. Rank-based techniques are regularly connected on the Web. In any case, Rank-based techniques are more troublesome to apply to email on the grounds that messages are practically continuously requested by time.

c) Prevention based method

In prevention based method [23] makes the spams contribution difficult. There is two type of prevention-based method; Interface based and Limit based. In Interface based method the interface restrict or deny access to the action that requires the users to change the contents. Some examples of the interface-based methods are *CAPTCHA*s for programmatic access to the websites, disposable email addresses for the programmatic access to the mails. In Limit-based method the numbers of resources required for the users to complete the action is usually limited. For example, the user can type the incorrect password three times before the accounts get blocked and later retrieve the passwords using the secret questions.

By using these methods we can easily reduce the spams and the phishing. Defence against the phishing can be mainly done on the user side. By using the password hashing the password that the user enters can be hashed so that the intruder cannot retrieve the password even the OSN user is using the same password for multiple sites. Another defence mechanism that can be proposed is to use the digitally signed emails so that we can reduce the amount of spoofing that will happen. This is mainly necessary because the users of the OSN often tend to share the personal information on their profiles.

7.2. Malware Attacks

The attackers along with the spams and phishing are also using malware attack to spread malicious entity to the OSN user space. Malicious entity can be worms, Trojan horses, viruses that has the capability to corrupt the user systems. Different malware attacks are described below.

a) Cross site scripting (XSS)

XSS attack [24] is a type of attack where the attacker runs part of the malicious code on the server. There are two types of XSS attack. Persistent XSS where the attackers code is permanently stored in the server HTML Text or as a message post. When the user visits the malicious code the attacker retrieves the information from the browser. In Non-persistent XSS attack the injected malicious code is sent back to the victim once he visits the server.

XSS worms affect the OSN in two steps. In the first step the attacker adds the malicious code to the victims profile such as link. Thereafter any person who visits this users profile will be injected with the malicious code to his profile by exploiting the AJAX flaws thus making the visitors profile infectious which helps the worms propagate.

b) Trojan worm

One of the popular Trojan worms [25] that came into light during the year 2008 was Koobface [26]. This worms propagates itself in the OSN and then send the messages using the deceiving topics to the users of the OSN. The users when click on the message, redirected to a third party website where they prompted to install the latest version of the flash player. Once the users downloads and installs the file, their system become converted as botnet and messages are sent from the infected users profile without the user being aware of the messages being sent to his friends making the worm spread all over the OSN. This worm was first identified on Facebook.

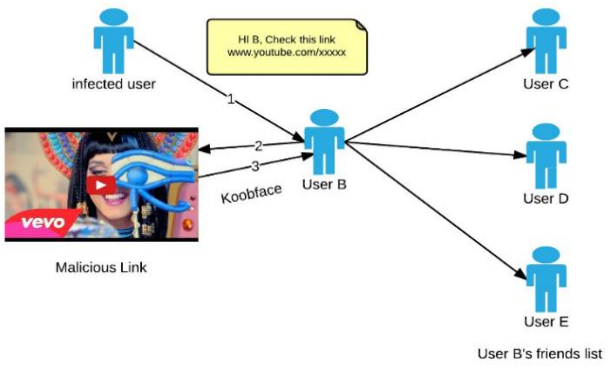

Figure 6. Koobface attack pattern [26]

7.2.1. Defensive Measures against Malware Attacks

Most prominent solution by researchers is to having a decoy friend which attaches to the set of the OSN users [27]. Decoys receive the evidence of the worms, the system performs correlation of local and network to distinguish between the normal user communication and the worm evidence.

7.3. Reverse Social Engineering Attacks

One of the most happening, vulnerable but still less popular OSN attack is Reverse Social Engineering Attack (RSE). The RSE attack mounted in two steps. In the first step the attacker wait the users to increase their curiosity. In the second step the attacker waits for the victim to contact and then exploits.

RSE is even more dangerous in the social networks. The first reason is the attacker can reach numerous of the victims those registered with OSN. Secondly, RSE can bypass the filter-based technique that will filter of the unrecognised contacts [28]. The third reason being that in RSE the attacker makes the victim to contact them, which make the less alarms to raise and making the attack to be successful.

7.3.1 Defence Measures Against Reverse Social Engineering Attacks

In [28] the authors proposed three solutions for countering the RSE attacks. First solution is for the friend recommendation. This feature is very useful in the social networks but the attackers can easily influence the OSN system to recommend the untrusted entity as their friends. Therefore, it very much necessary that the service provider show the friend recommendation only, if there is a strong connection between the two users. For example, a simple email lookup should not provide the friend recommendation but it should also consider many other factors such as their interests, common friends so on.

Second approach is by the use of RSE honeypot account, which receives the request from the other users but does not send the request. Thus identifying the fake account. This will happen when the new users join the OSN and sent the request and receive the requests. When searching the friend recommendation can be provided which is untrusted.

The third approach is to use *CAPTCHA*, which can be embedded with the incoming friend request. Since in the present OSN is full of the spams and phishing, the service provider can display with *CAPTCHA* before the friend's request are sent to others. But the present day OSN does not have any such protective measures. By introducing the *CAPTCHA* for accepting the suspicious friend request it increases the RSE attack difficulty to implement.

7.4. Distributed Denial of Service (DDoS) Attack

The present social networks such as Facebook, Twitter are made up of PHP, JavaScript, ASP.net, C, C++ etc., to exploit the DDoS attack the attacker has to embed the malicious application in the URI which will be linked to the victim server. These URI can be embedded in the user data such as the images, message, and notification. Whenever the OSN user interacts with the malicious application it produces the unsolicited HTTP Get request since the application is embedded in the social network it may trigger from OSN only but practically this trigger can also be raised from the targeted user's web browser. This, hence, make the series of request to be raised causing the denial of service attack.

7.4.1 Defence Measures against DDoS Attacks

Since the DDoS attacker send the request continuously to the server making the legitimate requests from not reaching the server hence make the server unavailable for legitimate users. Traffic filtration, traceback mechanism after DDoS occurred and many other solutions are existing to mitigate the attacks.

One approach that can be used to provide the defence is to use the decentralised architecture so the server do not congested by the requests. Also to provide the security in the client side that the users of the OSN should be aware of the applications content and authenticity can also decrease the frequency of such attacks to occur. . The users must have detective mechanism to beat DDoS attack. Various anomaly detection mechanism and existing traceback methods are being used to identify compromised bots are described in [29].

7.5. Sybil Attack

In Sybil attack [30] the adversary masquerades as the multiple users pretends to be multiple nodes on the network. If there exists a central authority it very easy to mitigate the Sybil attack. But in the decentralised architecture such as P2P there is no governing central authority. Present days P2P architecture try to mitigate the Sybil attack by identity and the IP address binding [29]. But the only drawback that exists in this method is that the attacker can easily obtain the IP addresses of the user. In the current state of Internet Flow or IP flow, it is highly likely to obtain the IP address of a legitimate user and transform that address i.e. spoof IP address or as the trusted users (but fake or compromised). Such compromised bots are used to send the requests to the server or affecting the other users.

Academicians particularly treat it as node formation. These botnets are nothing but the OSN user but the requests can be raised without the user being aware of it.

7.5.1 Defence Mechanism against Sybil Attack

We have surveyed that two effective defence mechanism discussed below are currently being used by Social network providers to provide defences against particular Sybil attack [2].

a) Trusted certification

When using this approach only the trusted entity [29] can enter into the network. This might be the only technique that can help the OSN providers to completely mitigate the Sybil attack problem. This technique requires a central authority having global view to verify each entering users. The limitation of this technique is an issue related to placement of centralised authority in the proposed decentralised architecture. However a centralised authority can be deployed in existing centralised architecture thus could makes this technique feasible. Instead of this, critical scalability issue may arise because of manual verification process.

b) Recurring costs

As explained above the Sybil attack cannot be launched until the Sybil nodes are formed. There are many approaches that exist that impose the cost when creating the nodes. One such example is that *CAPTCHA* which will increase the cost of node creation. *CAPTCHA* may help to reduce the requests raised from the botnets.

c) Resource testing

In this defence measure all the resources are tested such as bandwidth, network structure, storage ability and the IP address associated with node that representing the legitimate user must be investigated. In [31] the researchers have provided an optimal technique for the defence against the Sybil attack it is called "Sybil guard". Sybil guard is protocol that limits untrusted influence of the Sybil attack. This protocol is based on the social relationship of the user in the social network where an edge on the social graph between two users represents human established trust relationship. The adversary can create many Sybil nodes but only few relationships. Therefore there is a small cut between the honest nodes and the Sybil nodes. This protocol exploits the above-mentioned property that bounds the number of the Sybil nodes that the adversary can create.

8. Conclusion

Online Social Networks provide the new means for communication and interaction between the users of the OSN and is predicted to be two fold rise in users in a year. Unfortunately, because of flaws in existing functionality

and the architecture in OSN, vulnerabilities are studied and presented in this paper. We have also discussed the risks associated with the present OSN in this paper. In this work, the breaches happens in the OSN from the other users, service providers and also from the third party applications are briefly highlighted. Significantly we have highlighted the various attacks and there defence mechanisms in detail in this survey.

References

[1] Zhang, C.; Sun, J.; Zhu, X.; Fang, Y. (2010) Privacy and security for online social networks: challenges and opportunities. *Network, IEEE*, **24**(4): pp. 13-18.

[2] WEN, S; JIANG J; XIANG X.; YU, S.; ZHOU, W. (2014) Are the popular users always important for information dissemination in online social networks? : *Network, IEEE*: vol.28 (5): pp.64-67.

[3] WEN, S.; HAGHIGHI, M.; CHEN, C.; XIANG, Y.; ZHOU, W.; JIA, W., (2014) A Sword with Two Edges: Propagation Studies on Both Positive and Negative Information in Online Social Networks: accepted *in IEEE Transactions on Computers.*.

[4] JIANG J.; HUNG C.; WU J. (2010) Bandwidth- and Latency-Aware Peer-to-Peer Instant Friend cast for Online Social Networks, *IEEE 16th International Conference on Parallel and Distributed Systems (ICPADS)*, pp.829-834, Shanghai, 8-10 Dec. 2010

[5] SHARMA, R.; DATTA, A. (2012) SuperNova: Super-peers based architecture for decentralized online social networks: *Fourth International Conference on Communication Systems and Networks (COMSNETS)*, pp.1-10, Bangalore 3-7 Jan. 2012.

[6] MEI, Y.; WANG, J.; WANG, Q.; LIU, X.; ZHAO, Z.; ZHANG, Y. (2010) Semantic-based query routing in P2P social networks: *IEEE International Conference on Wireless Communications, Networking and Information Security (WCNIS)*, pp.674-678, Beijing, China, 25-27 June 2010.

[7] XIONG, Z.; JIANG, W.; WANG, G., (2012) Evaluating User Community Influence in Online Social Networks: *IEEE 11th International Conference on Trust, Security and Privacy in Computing and Communications (TrustCom)*, pp.640-647, Liverpool, 25-27 June 2012

[8] SHI, L.; YU, S.; LOU, W.; HOU, Y.T. (2013) SybilShield: An agent-aided social network-based Sybil defence among multiple communities: *INFOCOM, 2013 Proceedings IEEE*, pp.1034-1042, Turin, 14-19 April 2013.

[9] SADEGHIAN, A.; ZAMANI, M.; SHANMUGAM, B. (2013) Security Threats in Online Social Networks," Informatics and Creative Multimedia (ICICM), *International Conference on*, pp.254-258, Kuala Lumpur 4-6 Sept. 2013.

[10] ALSERHANI, F.; AKHLAQ, M.; AWAN, I.U.; CULLEN, A.J.; MIRCHANDANI, P. (2010) MARS: Multi-stage Attack Recognition System: *24th IEEE International Conference on Advanced Information Networking and Applications (AINA)*, pp.753-759, Perth, WA, 20-23 April 2010.

[11] GAO, H., HU, J.; HUANG,T.; WANG, J.; (2011) Security Issues in Online Social Networks. *IEEE Internet Computing*, vol. 15(4): p. 56-63.

[12] NARAYANAN, A. AND V. SHMATIKOV. De-anonymizing Social Networks (2009). in *Security and Privacy, 30th IEEE Symposium*.

[13] JAE YOUNG CHOI; DE NEVE, W.; PLATANIOTIS, K.N.; YONG MAN RO (2011) Collaborative Face Recognition for Improved Face Annotation in Personal Photo Collections Shared on Online Social Networks,", *IEEE Transactions on Multimedia*, vol.13, no.1, pp.14,28.

[14] DEB, S, ZHANG Y. (2004) an overview of content-based image retrieval techniques: *18th International Conference on Advanced Information Networking and Applications, AINA 2004*, vol.1, no., pp.59, 64.

[15] BAI Y., SU; X. BHARGAVA, B. (2009) Detection and filtering Spam over Internet Telephony — a user-behaviour-aware intermediate-network-based approach: *IEEE International Conference on Multimedia and Expo. ICME 2009*, pp.726, 729. NY, June 28 2009-July 3 2009

[16] WONG, K.; WONG, A.; YEUNG, A.; WEI FAN; SU-KIT TANG (2014) Trust and Privacy Exploitation in Online Social Networks," *IT Professional* , vol.16, no.5, pp.28,33.

[17] ERLANDSSON, F.; BOLDT, M.; JOHNSON, H. (2012) Privacy Threats Related to User Profiling in Online Social Networks: *International Conference on Privacy, Security, Risk and Trust (PASSAT), 2012 and International Conference on Social Computing (SocialCom)* , pp.838-842. Amsterdam, 3-5 Sept. 2012

[18] BODRIAGOV, O.; KREITZ, G.; BUCHEGGER, S. (2014) Access control in decentralized online social networks: Applying a policy-hiding cryptographic scheme and evaluating its performance: *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops)*, pp.622- 628, Budapest 24-28 March 2014

[19] Singh, K.; Bhola, S.; LEE, W. (2009) XBook: Redesigning privacy control in social networking platform: proceedings of 18th Conference of USENIX symposium, pp,249-266

[20] Symantec. *Spam and Fraud Activity Trends*. 2011; http://www.symantec.com/threatreport/topic.jsp?id=spam_fraud_activity_trends&aid=analysis_of_spam_activity_trends

[21] GULYAS, G.G.; IMRE, S. (2014): Measuring importance of seeding for structural de-Anonymization attacks in social networks: *IEEE International Conference on Pervasive Computing and Communications Workshops (PERCOM Workshops), 2014* , pp.610-615, Budapest 24-28 March 2014

[22] HUBER, M.; MULAZZANI, M.; WEIPPL, E.; KITZLER, G.; GOLUCH, S. (2011) Friend-in-the-Middle Attacks: Exploiting Social Networking Sites for Spam: *Internet Computing, IEEE*, vol.15, no.3, pp.28-34.

[23] DING X.; ZHANG L.; WAN Z; GU. M. (2010) A Brief Survey on De-Anonymization Attacks in Online Social Networks: *International Conference on Computational Aspects of Social Networks (CASoN), 2010*, pp.611-615, Taiyuan 26-28 Sept. 2010

[24] JOHNS, M.; ENGELMANN, B.; POSEGGA, J. (2008) XSSDS: Server-Side Detection of Cross-Site Scripting Attacks: *Computer Security Applications Conference, 2008. ACSAC 2008. Annual*, pp.335-344, Anaheim, CA, 8-12 Dec. 2008

[25] FAGHANI, M.R.; MATRAWY, A.; CHUNG-HORNG LUNG (2012) A Study of Trojan Propagation in Online Social Networks: *5th International Conference on New Technologies, Mobility and Security (NTMS), 2012*, pp.1-5, Istanbul , 7-10 May 2012

[26] THOMAS, K.; NICOL, D.M (2010) The Koobface botnet and the rise of social malware: *5th International Conference on Malicious and Unwanted Software (MALWARE), 2010*, pp.63-70, Nancy, Lorraine, 19-20 Oct. 2010

[27] WEI XU, F.Z., ZHU S. (2010) Toward Worm Detection in Online Social Networks. P*roceddings of 26th Annual computer security application confrence, 2010, PP. 11-20* Austin, USA,December 06 - 10, 2010.

[28] IRANI D.,BALDUZI M. BALZAROTTI M, KIRDA E, PU C. (2011), pp 55-74 , Reverse Social Engineering Attacks in Online Social Networks: *"Detection of Intrusions and Malware, and Vulnerability Assessment*", Springer.

[29] YU.S. (2014) Defence against DDoS Attack**s**, Yu S**.,** *"Distributed Denial of Service Attack and Defence*", Springer.

[30] TRIFA, Z.; KHEMAKHEM, M..; (2012) Mitigation of Sybil Attacks in Structured P2P Overlay Networks: *Eighth International Conference on Semantics, Knowledge and Grids (SKG), 2012*, pp.245-248, Beijing, 22-24 Oct. 2012

[31] HAIFENG YU; KAMINSKY, M.; GIBBONS, P.B.; FLAXMAN, A.D. (2008) SybilGuard: Defending Against Sybil Attacks via Social Networks: *IEEE/ACM Transactions on Networking*, vol.16 (3), pp. 576-589.