SOCIETY FOR SCIENCE AND EDUCATION
UNITED KINGDOM

TNC Transactions on Networks and Communications

VOLUME 3, ISSUE 2
ISSN: 2054 - 7420

A Comparative Analysis of Privacy Preserving Techniques in Online Social Networks

[1]Firdous Kausar and [2]Shoroq Odah Al Beladi
Department of Computer Science, College of Computer and Information Sciences,
Al Imam Mohammad ibn Islamic Saud University, Saudi Arabia
[1]firdous.kausar@ccis.imamu.edu.sa, [2]shorogodah@gmail.com

ABSTRACT

The world became a global village due to the great revolution in communication and network fields. The internet eases the communication process all over the world. The social network services, for example; Facebook, that occurred recently are considered one of the most essential and common outcomes due to this revolution. People from different ages, views, cultures, languages, religions, education levels, etc. from all over the world can easily communicate through the social networks. Furthermore; people can share their news and personal information with each other. It can be also noticed that the social network are extensively used, so the privacy issues within it is essentially to be considered. This privacy will prevent any illegal accessing for the user personal information which in turns will increase the users' conviction to use these networks during their daily life and encourage the other people who have never use the services that are available for using by social networks. This paper analyses several techniques which provides the privacy for Online Social Networks including Reclaim, Safebook, K-Automorphism, Vis-à-vis and SPKA. These techniques are investigated and clarified in terms of their methods in addition to the advantages and drawbacks.

Keywords: Online Social Networks (OSNs), Privacy, Communication and Transport (CT), Social Networking Services (SNSs), Distributed Hash Table (DHT), Virtual Individual Server (VIS), Trusted Identification Service (TIS).

1 Introduction

The great revolution that emerged recently because of lead to a clear development in all life filed. This development is required to enable the users all over the world performing their tasks and works in addition to enhance the communication tools and skills among the users all around the world. Due to this revolution; "Online Social Networks (OSNs)", including Facebook, Twitter and Google +, emerged as a common and deployed communication tools that can be used for sharing the information between these users. These networks have attracted a very large number of users. Furthermore; the adoption is still increasing day by day [1].

Several services are provided to the users due to using OSNs, such as; instant messages, internet phones and blogging without giving considerations for the physical location of the users. Furthermore; the friends and families could easily maintain their relations in more reliable and convenient manner in comparison with old style contact techniques, such as; phone conversation and emails. These networks in turns also increase the popularity for the users within social groups; all these are considered attractive advantages for OSNs. On the other hand; there is some drawbacks for OSNs that limits the deployment of it and always increase the motivation for new techniques

DOI: 10.14738/tnc.32.854
Publication Date: 20th April, 2015
URL: http://dx.doi.org/10.14738/tnc.32.854

that will shrink and eliminate the effect of these drawbacks, the privacy concern is considered one of the most critical and significant issue within OSNs. The privacy risks increases in case that the personal information for the users, such as names; are included within OSNs applications. As a result; there is an increasing demand for novel privacy schemes that will in turns preserve the users' personal information and ensure their privacy [2].

The privacy is considered one of the OSNs security objectives in addition to the availability and integrity [12]. Maintain the privacy of the users is the most essential and critical objective for "Social Networking Services (SNSs)". The privacy issue is concerned in personal information protection that is shared and published on the profiles of users in addition to maintaining the privacy during communication [11]. So; only trusted parties are able to trace the communicating parties. In addition to that; the privacy is also concerned in hiding the message details in such way that only the receiver and sender can recognize it. To conclude all; the privacy is concerned in personal information hiding for the users. The privacy should be achieved by default [3].

The breaches of privacy within social network can be classified into three main groups, which are [4];

- Identity detection; this type of breach occurred in case that individuals that the record belong to is disclosed. This will in turns result in information revelation of users and the shared relationship from them with other network's individuals.
- Sensitive link detection; this type of breach occurred in case of the association revelation among two users. This information is generated by social activities in case of utilizing the services of social media by the users.
- Sensitive attribute detection; this type occurred in case that the confidential and sensitive user information is being attained by the attacker. Sensitive attributes are related to the link and entity relationship.

Three levels are included within SNSs as illustrated below in Figure 1 below [3].

Three levels are included within SNSs as illustrated below in Figure 1 below [3].

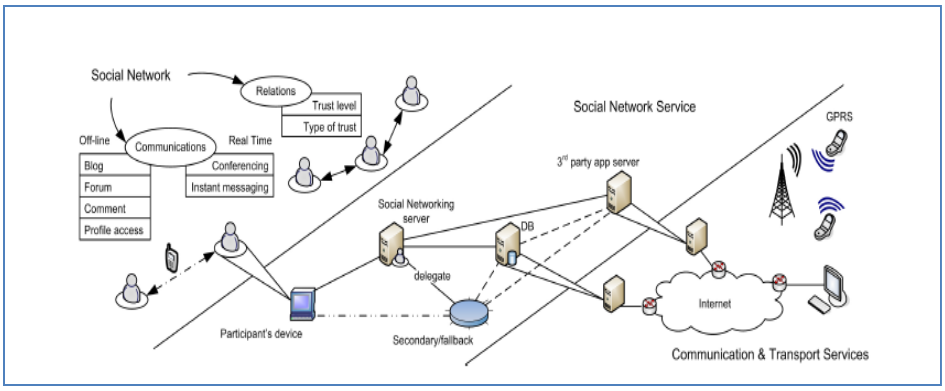


Figure 1: levels of SNSs [3]

As illustrated in figure 1; the SNSs levels can be summarized as listed below;

- The level of "Social Network (SN); this level includes the members relationship digital representation.
- The level of SNS; in this level; the infrastructure of the application that is controlled through SNS provider.
- The level of "Communication and Transport (CT)", this level includes the transport and communication services that are provided through the network.

During SN level; the members are provided with several functions regarding to the real life interaction, such as; profile accessing, like, finding friends and commenting. All these functions are implemented relying on the level of SNS. SNS level in turns includes the main services that are required or needed for generating SN services, including, storage, web services and communication. The delegation and redundancy are considered two of the most commonly deployed techniques for availability improvement [3].

Data retrieval and storage, content indexing, data access permission management in addition to the leave or join of node, are all implemented using either decentralized or centralized fashion of distribution within SNS level. Furthermore; the internetworking infrastructures and protocols that were previously implemented within CT level are used during SNS level. Depending on OSN architecture; the attacker can be defined as of the following type [3];

- SN level malicious member
- SNS level service provider.
- CT level malicious that have illegal access for the infrastructure.

One main aspect that must be achieved within OSNs is the privacy issue; it is mainly concerned in protecting the identity of the members. For example, the identity theft is defined as malicious service provider or member obtains the authorized users credentials and then take action rather than them by getting access for the profiles of these members. Furthermore; the intrinsic trust among people also plays a great role in privacy violation for the users via making another copy of the target personal profile using the personal information and then start communicate with others via sending them friend request for example. Another type of probable attacks is the profile porting, in this type; the target person profile is created in OSN by the attacker without present of the victim. The detection for this type of attack is usually difficult mission [3].

Several research papers and studies investigated and evaluated the privacy issue within OSNs. Generally; the privacy can be achieved through network decentralization or encrypting the data before storing it. The stimulants of reading the private data is the condition for sustain the centralized network. In decentralized type of network, there is no incentive that the network relies on, so; the operation of the network can be continued depending on the contributed resources of the user [5].

2 Privacy Preserving Techniques

In this section we provide the description of different privacy preserving techniques.

2.1 Cachet

The cachet [1] is a structural design by which a strong privacy and security can be achieved for the users maintaining the OSNs main functionality. The availability, confidentiality and integrity in addition to the users' relationship privacy can be protected using cachet. The user data is stored using distributed nodes pool; this pool is also used for availability ensuring [14]. Since the cachet storage nodes are not trusted; then a technique of leverage cryptographic, specifically; "Attribute-Based Encryption (ABE)", is employed in order to achieve the trust for the storage nodes within cachet, which in turns will protect the data confidentiality [13]. Furthermore; "Hybrid Structured-Unstructured Overlay Paradigm (HSUP)" is also employed to achieve efficient retrieval and dissemination for the data. So; a "Distributed Hash Table (DHT)" is augmented regarding to the users social links. In order to minimize the overhead on the network, for example; reducing cryptographic, then the recent updates within SN is stored by the social contacts that operate as caches. FreePastry

Simulator was employed in implementing the cachet prototype. Furthermore; newsfeed application was also implemented to illustrate the existing OSNs functionality. An example for this technique is illustrated below in the following figure 2

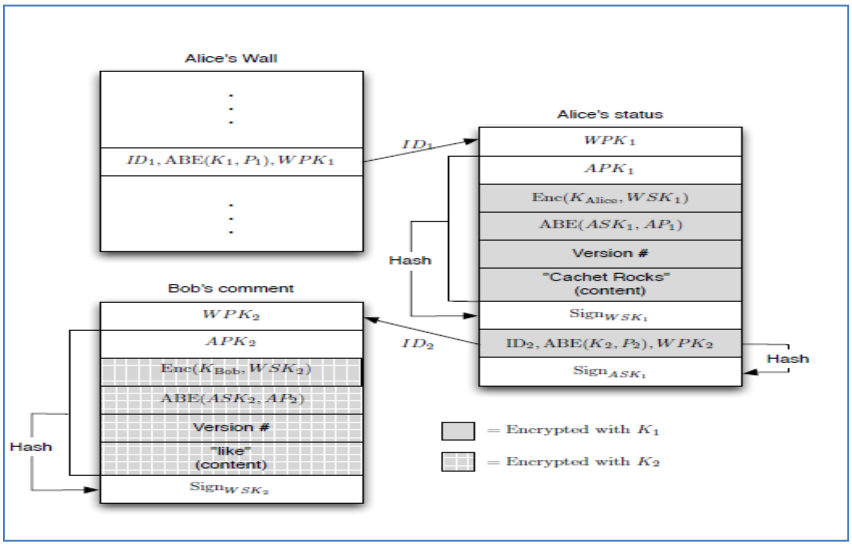


Figure 2: Cachet Example [1]

For Alice to be joined to the cachet; the varouis keys should be firstly generated, a wall and profile inforamation is created and finally these inforamation are saved as wall and root information within DHT respectively. In order to set up a friend or co-worker relationship with Bob, then an a secret key is generated for Bob using ABE among the co-worker and friend attributes. A different realtionship may be established by Bob among Alice. If Alice wants to post a status update, then a status object is created by alice, in addittion to the number of version, content and secret and public keys for appended polices and write (WPK1,WSK1,APK1, ASK1). “Write Ploicy Sgnature Key (WSK1)” is used in signitaure genreation. A symettric key for encryption say K1 is randomly picked by Alice and use it in encrypting the objects (excluding APK1, WPK1 and signature) [1].

Furthermore; an ID say ID1 is also selected randomly and then used in inserting the objects within DHT. A refreence is finally created for the update of the status including K1, ID1, WPK1 and then added to the wall. This update can be read by Bob through finding referrence on the wall of Alice, after that the attribute-based key that was prevouisly transmitted by Alice for alice will be used in decrypting K1. The object is then retrieved from DHT in additiion to ID1, the encrypted field is then decrypted employing K1. The object integrity is then being ensured through signature verification. If Bob wants to comment on the status update of Alice; then the same steps that were followed by Alice during creating the update will be followed in creating the comment object. Append operation is then used in referrence inserting realted to the novel object into the update of Alice. If AP1 is staisfied, then ASK1 is decrypted to be latterly employed in signature geneartaion.

2.2 ReClaim

ReClaim[5] is a solution for decntralized OSNs, each one of the peers is named as Reclaim. friends of friend are emplyed in replicating the data, so; no addittional server is required for storing the data. The public keys are exchanged in order to set up the frindship, these keys are in tuns employed in encrypting all messages and make them redable for only the target user. “Private-Set-Intersection (PSI)” approch [10] is employed in the current online friends. So; two peers are allowed to discover the common friends without disclosing the unmutual ones. Missed or older messages are then syncronized to common friends by the two peers using Bloom filters [9]. the encryption for all messages that were prepared for the peers and their friends can be performed effciently, furthermore; the duplicate transfers of messages is prevented. By this way; the operation of ReClaim can be reliably continued despite of intermittent connection of the network, the firewall of "Network Address Translation (NAT)" and network dealy. A method that is recognized as $FSF_{A,B}$ is employed in detrmining whether that the just connected peers are either friend or hanve common friends. The SFS inputs is denoted by Ƒ and it is recognized as friendset that includes the user and the friends identifiers.

2.3 Vis-à-vis

The concept for this decentralized framework is summarized in maintaining the privacy for the "Virtual Individual Server (VIS)". In this technique, each person data is stored within him/her VIS. Vis-à-vis is a privacy technique that is concerned in location information privacy. The information location can be efficiently shared within groups through employing the trees of distributed location. The architecture for Vis-à-vis is illustrated below in the following figure 3.

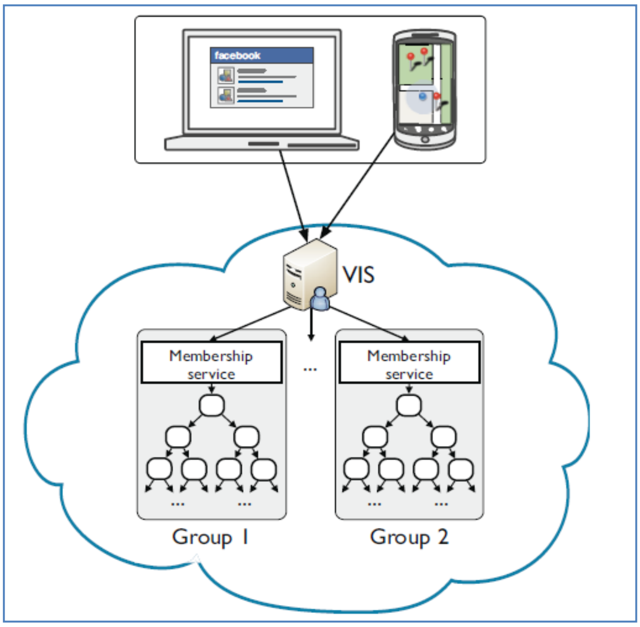


Figure 3: architecture of Vis-à-Vis [6]

The process of naming the groups is performed through employing a descriptor which in turns includes the public key related to the group in addition to specific string used to transfer the shared attribute between the members of the group. The expression for the descriptor is given by$< K +_{owner}, string >$; $K +_{owner}$ represents public key. No sensitive information in terms of the privacy is included within descriptors, for instance; the public key for the user is shared among the group. As shown in Figure 3; the architecture of Vis-à-Vis looks like the structure of distributed tree. The hierarchical structure is adopted to be used over DHT; since the required range quires for search operation is not easily provided when applying DHT. The groups that based on the location are accessed by the users throughout clients, for examples; web browsers and mobile applications. Vis-à-vis was designed to deal with established OSNs, for example; Facebook. A stronger privacy is achieved for the users when employing Vis-à-Vis in comparison with centralized services, for examples; MySpace and Facebook. This happens because the users are provided with more control for their personal data and who can access it. The trust model is designed based on the compute utilities business interest in addition to the user's social relationships [6].

2.4 Safebook

Safebook is an architecture in which three tiers are included in addition to the layers direct mapping with OSNs level that were previously introduced, this architecture is illustrated below in the following figure 4.

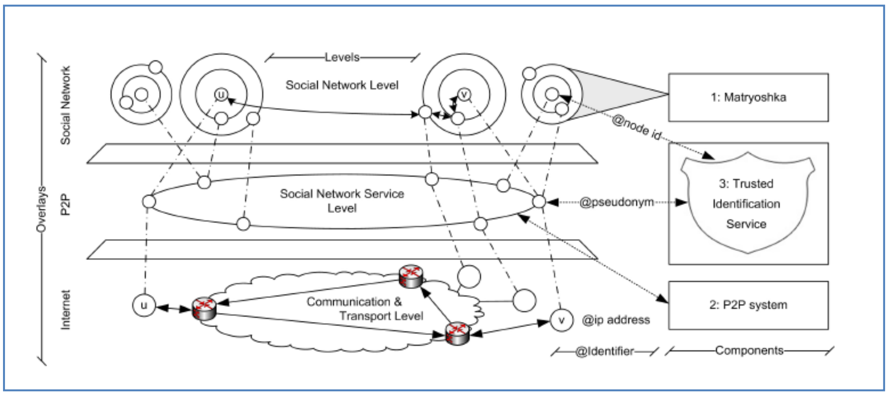


Figure 4: Safebook architecture [3]

From Figure 4; it can be concluded that;

- CT level is represented by Internet.
- SNS service is implemented using “Peer to Peer (P2P)” base.
- SN level is implemented using user-centered layer of social network.

A node is used to present Each Safebook party. In internet, this node is host node, in case of P2P overlay then it is peer node and member within SN layer. Two types of overlay are formed by the Safebook nodes, which are [3];

- P2P base or substrate; this type is concerned in lookup service providing.
- Matryoshkas set; this type is considered SN concentric structure and it concerned in data storage providing in addition to the creation of communication privacy around all nodes. The structure of Matryoshkas is illustrated below in the figure 5.

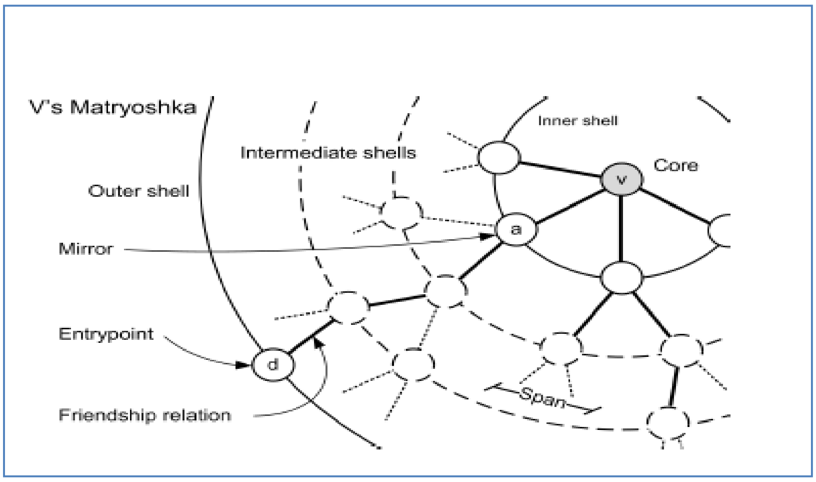


Figure 5: Structure of Matryoshkas [3]

Furthermore; a “Trusted Identification Service (TIS)” is alfo featured by safebook in order to give an unambigous identified for each node; the Pseudonym in addittion to the “Node Identifier (NI)”. Particular countermeasures set is implemented by each component in Safebook in order to overcome the probable threats in OSNs. As shown in Figure 5; each single Matryoshkas is concerned in providing the protction for the core node that is addressed using NI within SN layer. Radial paths are used to connect the nodes; so the message can be recursively transmitted among shells. The trust relationship that is the most similar to social network is considered by the paths; so a hop is used to connect two nodes related to users who have real-life trust relationship. The mirrors are the nodes that have direct contact with core; the data are stored in these nodes in encrypted form. The remaining nodes excluding the core and mirrors are recognized as entrypoints which in turns act as gateway for passing all request of data to the core node [3].

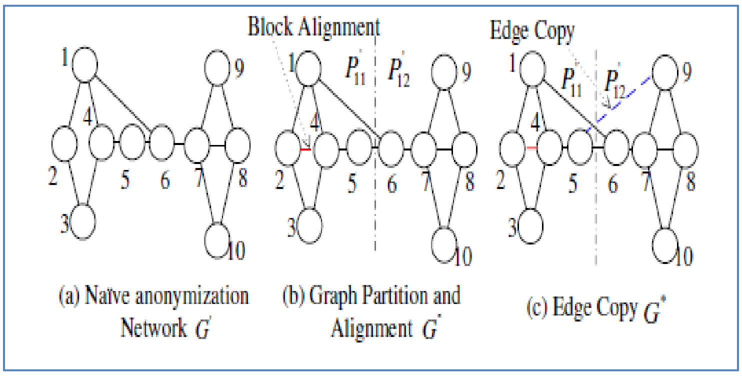
2.5 K-Automorphism

K-Automorphism is also a privacy technique for OSNs; it is concerned in the problem of identity disclosure through employing "K-Match (KM)" technique. Table 1 summarizes illustration for the main symbols that are employed within this algorithm [7].

Table 1: K-Automorphism symbols [7]

Symbol	Description
G	Main network
$G^{´}$	Gullible anonymized network
$G^{´´}$	Anonymized graph with alignment and portion
G^{*}	Anonymized graph with KM technique
$\overline{G^{*}{}_{\iota}}$	Anonymized graph with GenID technique at T_t time.
U_i	Blocks group.
P_{ij}	A single block within U_i group

The privacy within K-Automorphism is provided against structural attack as summarized as follow; for network G; if a $k-1$ numbers of automorphic functions F_a $(a = 1, 2, \dots \dots, k-1)$ are defined within G, and the relation $F_{a2}(v) \neq F_{a2}(v)$ for each v where v is a vertex in G, then the K-automorphic network take place. The vertex v cannot be differentiated from other similar vertex depending on the structural information [7]. The privacy in this technique is achieved through using KM technique that is summarized below in the figure 6.



(a) Naïve anonymization Network $G^{'}$

(b) Graph Partition and Alignment $G^{''}$

(c) Edge Copy G^{*}

Figure 6: KM technique [7]

As shown in figure 6; if 2-different matches are required to be guaranteed within network G, then g is firstly divided into two blocks (Figure 6b). The graph alignment is then performed to attain two aligned blocks. In case that query Q match is included within G^{*}, then 2 matches are least exist of Q within G^{*}. The principle of k-different match is still able to be applied after performing the alignment and partition for the network [7].

2.6 Single Pass K-means Anonymization (SPKA)

This technique mainly based on the method of K-mean clustering by including only one iteration rather than multiple ones. If D represents the records set that will be anonymized, K represents the clusters' number and n represent the records' number. The SPKA technique can be summarized as follow. The quasi-identifiers are used to classify the records that are within D. a K records then randomly selected to represent the early cluster that will used in generating all later K clusters. For $r \in D$; r is assigned to the center of nearest cluster. The centers of the clusters are updated in case of adding novel records. In order to decrease the losses of information, some adjustments are then performed. Some records are taken away from the clusters that consist of number of records that is larger than k and then added to these clusters that consist of number of records that is less than k. in case of not having clusters with number of records less than k; the removed records are then allocated within particular closest clusters. As a result; a complexity of $O(n^2/k)$ is achieved. The t-closeness privacy measure was employed during their investigation. In this technique; the distance among the sensitive attribute distributions and the total table attribute distribution is required to be less than predefined threshold value. "Earth Mover's Distance (EMD)" is employed in computing the distance among two distributions.

3 Comparative Analysis

The privacy and security are considered two of the most essential and critical concerns in social network. Several techniques can be employed to achieved these two concerns, Six techniques were introduced and investigated during this research, which are; Reclaim, Cachet, Vis-à-vis, Safebook, K-Automorphism and SPKA. Each one of the considered technique has it is own concepts and methodology in order to provide the social network users with the required level of privacy, which in turns result in protect their personal information from being accessed by unauthorized persons.

The first considered technique is Reclaim technique; this technique is mainly based on employing PSI, $\mathrm{FSF}_{\mathrm{A,B}}$ method and Bloom Filters in order to provide achieve users' privacy. This technique has several advantages that make preferable to be used and employed in social network. The first point is the decentralized architecture; which in turns means that the communication can be performed between peers and users without need for any central equipment. Furthermore; this techniques has the advantages of low running costs, able to deal with extreme churn, overcoming NAT firewall, enabling the communication among the friends despite of being offline and accessing the replicas easily from the users. On the other hand; the complexity within Reclaim structure may limit the deployment for this technique.

The second considered technique is Cachet technique; which mainly implemented based on Nodes distributed pool, ABE, HSUP concepts in addition to Social contact employment. High level of security and privacy is provided to OSNs users. Furthermore, this technique provides the users with availability, confidentiality and integrity protection. On the contrary; this technique essentially still needs Pliability against the churn of node. Furthermore; there is difficulty in recognizing the peer to peer connection for the users who are located at the NAT back. This technique also required additional bandwidth, computational resources and data storage.

The third technique is Vis-à-vis technique. This technique based on VIS and Distributed trees concepts. Decentralized framework is also considered in this technique in addition to Hierarchical structure. This technique was designed to achieve privacy on location information only and some improvement and development is still required to be added to the design in order to include other information privacy. Furthermore; the breaches cannot be totally eliminated.

The fourth technique is the Safebook; which mainly based on Matryoshkas, DHT and Real-life trust concepts in order to provide the OSNs users with the required level of privacy and security. The decentralized structure is also considered here. Furthermore; Feasible Realistic compromise among performance and privacy is also provided. Extra costs are required in case of using additional hobs to enhance the privacy.

The fifth considered technique is K-Automorphism, which based on KM algorithm, Edge copy, Graph alignment and Graph Partitioning concepts. The structural attack can be got over using this technique. Another advantage for K-Automorphism is that No uncertainty for the released network. The privacy can be also provided in dynamic release case. The complexity in finding automorphic functions may limits the deployment for this technique.

The last technique is SPKA that is mainly based on Clustering and T-closeness approaches. Similarity attack can be avoided using this technique. It also achieves better privacy level in comparison with l-diversity technique. An additional benefit is the Quasi-identifier data prevention. The complex process and the added overhead for anonymization process are two limitations for SPKA technique.

The table2 introduces a comparison between the considered privacy techniques for OSNs. The comparison is held based on the main concepts that were employed during the implementation and investigation of the techniques in addition to the benefits that encourage the use of these techniques in OSNs services and applications. The drawbacks that limit the deployment of them are also concluded.

Table 2: Comparison between different techniques

Technique	Main concepts	Strength/Benefits	Weakness/Challenges
Reclaim	-PSI -$FSF_{A,B}$ method -Bloom Filters	-Decentralized architecture. -Very low running costs. -Ability to deal with extreme churn. -Get over NAT firewalls. -Enable the friend communicating despite of being offline. - Easy replicas accessing from the users.	-Complex Structure.
Cachet	-Nodes distributed pool. -ABE -HSUP Social contact employment.	-Strong privacy and security in addition to maintaining OSNs main functions. -provide availability, confidentiality and integrity protection in addition to preserving privacy. -Practical ABE_decryption computational overhead.	- Still essentially requires pliability against the churn of node. -there is a difficulty in recognizing the P2P connection for the users who are at the back of NAT. -Extra computational resources. - Volunteering bandwidth and data storage.

Vis-a-Vis	-VIS. -Distributed trees.	-Decentralized framework. -Hierarchical structure.	-supply the users with only location information Privacy and some improvement is still achieve the privacy for other data types. - The breaches cannot be totally eliminated
Safebook	-Matryoshkas. -DHT -Real-life trust	-Decentralized structure. - Feasible Realistic compromise among performance and privacy.	-Additional costs is needed for increasing privacy be adding additional hops.
K-Automorphism	-KM algorithm. -Edge copy -Graph alignment -Graph Partitioning	-Can overcome structural attack. -No uncertainty for the released network. -Provide Privacy in case of dynamic releases.	-Complex process for finding automorphic functions.
SPKA	-Clustering. -T-closeness	- Quasi-identifier data prevention. -Outperforms the privacy of l-diversity technique. - provide privacy against similarity attack.	-Overhead due to applying anonymization process. -Complex technique.

4 Conclusion

The OSNs are now widely deployed and used by large number of users from all ages and different views all over the world. The security within these networks is considered essential concern that must be achieved in order to maintain the personal information of users from being accessed or recognized by un-authenticated users or attackers. The privacy is one of the security objectives within OSNs in a line with the integrity and the availability. This paper introduced six common techniques used in preserving the privacy within OSNs. A brief explanation for the techniques was introduced in addition to investigation of the main concepts that were employed to implement the technique. As a result of comparative analysis between the different privacy preserving techniques it has been found that each technique has its own potential benefits and drawback. Some of the drawback of these techniques include 1) have complex architecture 2) require excessive computational recourse 3) need special hardware and 4) not simple to implement. On the other hand these also provide a strong benchmark for providing anonymity and privacy in online social networks.

REFERENCES

[1] Nilizadeh, S, Jahid, S and Mittal, P, "Cachet: A Decentralized Architecture for Privacy", the 8th ACM international Conference on emerging networking experiments and technology, 2012.

[2] Sun, J, Zhu¢Ó, X and Yuguang Fang, Y, "A Privacy-Preserving Scheme for Online Social Networks with Efficient Revocation", IEEE INFOCOM, 2010.

[3] Cutillo, L, Molva, R, Strufey, T and Eurécom, I, "Safebook: a Privacy Preserving Online Social Network Leveraging on Real-Life Trust", 2007.

[4] Vijayalakshmi, V, Arunachalam, A and Nandhakumar, R. "Mining Social Media-Utility Based Privacy", (IJCSIT) International Journal of Computer Science and Information Technologies, 5 (4), 5480-5485, 2014.

[5] Zeilemaker, N and Pouwelse, J. "ReClaim: a Privacy-Preserving Decentralized Social Network", 4th USENIX Workshop on Free and Open Communications on the Internet, August 2014.

[6] Shakimov, A, Lim, H, C´aceres, R, Cox, Li, P, Liu, D and Varshavsky, A. Vis-`a-Vis: Privacy-Preserving Online Social Networking via Virtual Individual Servers. IEEE, 2011.

[7] Zou, L, Chen, L and Ozsu, M, "KAutomorphism: A General Framework for Privacy Preserving Network Publication", 2009.

[8] Poulin,I and Kani, M, "Preserving the Privacy on Social Networks by Clustering Based Anonymization", International Journal of Advanced Research in Computer Science & Technology (IJARCST), 2 (1), pp.11-14, Jan-March 2014.

[9] Bloom B. H. Space/Time Trade-Offs nn Hash Coding with Allowable Errors, Communications of the ACM 13, pp. 422–426, July 1970.

[10] Freedman, M. J., Nissim, K., And Pinkas, B. Efficient Private Matching and Set Intersection, in EUROCRYPT '04, pp. 1–19, 2014.

[11] Beato, F., Kohlweiss, M., and Wouters, K, Scramble! Your Social Network Data, in PETS '11, vol. 6794 of Lecture Notes in Computer Science, pp. 211–225, 2011.

[12] Elena Z, Lise G, Privacy in Social Networks: A Survey, in Social Network Data Analytics, pp 277-306, 2011.

[13] J. Bethencourt, A. Sahai, and B. Waters. Ciphertext-policy attribute-based encryption. In IEEE Security & Privacy, 2007.

[14] S. Jahid, S. Nilizadeh, P. Mittal, N. Borisov, and A. Kapadia. DECENT: A decentralized architecture for enforcing privacy in online social networks. In SESOC, 2012.