JETIR.ORG ISSN: 2349-5162 | ESTD Year : 2014 | Monthly Issue
JOURNAL OF EMERGING TECHNOLOGIES AND INNOVATIVE RESEARCH (JETIR)
An International Scholarly Open Access, Peer-reviewed, Refereed Journal

Data confidentiality Algorithms for Online Social Networks

Mrs. Rachana Ashtekar [#1]**, Dr. A. V. Deshpande** [*2]

[#1] *Department of Computer Engineering, Smt. Kashibai Navale college of Engineering, Pune (Affiliated to SPPU- Pune)*

[#2] *Department of Computer Engineering, Smt. Kashibai Navale college of Engineering, Pune (Affiliated to SPPU- Pune)*

Abstract— **Social networks, which are practically a part of our daily lives, have created new social norms for communication and behavior. Even though people and businesses have been using social networks extensively for years, governments are becoming more and more interested in the latest in communication technologies. Sites like Facebook, Twitter, and Linked In offer a way for people to connect based on various things like already-existing friendships, shared hobbies, or employment. . Its openness increases the likelihood of vulnerabilities, data breaches, and compliance violations. It also creates additional opportunities for hackers to flourish due to a lack of regulation and standardization, making it the most recent platform for security attacks. For security reasons, it is required to provide some data confidentiality algorithm for online social network to solve this problem. This paper provide the comprehensive review of social network security threats and existing solution that can provide the security for the social network user. The proposed model provide the functionality like, revealing hidden attribute value of social profile, node similarity, privacy- preserving social network analysis and privacy-aware access control. In this paper the mainly uses the approach to privacy as a protection model, structural model evaluation, OIGH algorithm. The proposed schema used extended-OIGH method contain the fednoise algorithm and split and carry algorithm.**

Keywords— **Online Social Networks, Woman business support network, Attribute management server, Privacy Enhancing Technologies, Social Networking Sites (SNSs).**

I. INTRODUCTION

Online social networking is gaining popularity. With 2.45 billion monthly active users as of September 2019, Facebook is the largest social networking service provider in the world. Social networking websites have, on the one hand, drastically changed how people communicate with one another and meet new people [1]. However, there are a lot of privacy issues that have been brought up by social networking websites. For instance, Facebook has been charged with disclosing user information to firms like Apple, Spotify, Netflix, Microsoft, and Yandex in addition to online merchants like Amazon and Apple. According to Facebook, Cambridge Analytica inappropriately stole personal information from millions of Facebook users for political advertising. One of today's most widely used venues for online communication is online social networks (OSN) [4], [5]. Due to technology innovation, everyone today has simple access to this communication platform. Our platform generates enormous volumes of data each day, including confidential user information. The only issue is the security of the user data. Sensitive information is stored by OSN providers in decentralized locations, where it is divided among a large number of servers and is so within the control of the server owner [3]. The data is therefore vulnerable to hacking and data theft. There is little user control over how their data is used. The primary source of income for the OSN is the commercialization and monetization of the data. The OSN analyzes the data to add or remove services based on user interest. The OSN emphasizes networking through online relationships, correspondence, and sharing. OSN providers are quickly incorporating the 6 recommendation system into their products to replace buddy referrals as the primary way to find new friends [2].

The paper gives a technical overview of how OSN data is encrypted and protected using homomorphic encryption and trust calculation. Social networking services (SNS) are a major part of the Internet. They provide a wide range of services that are intended for a large user base with different social, educational, and national backgrounds. They also enable communication even for those with little technological know-how. In general, online social networking (OSN) [7]. The results of these SNSs are digital representations of the subset of relationships that their participants, registrants, or institutions entertain in the real world. Through their relationships with participants, they model social networks as graph social networking services as messaging and social Platform not only attract loyal users trying to add value to the community, but also parties with considerable adverse interests, whether they are commercial or malicious, and the main motivation for members to join OSN, create profiles and use the service [6], [8]. Different applications available to easily share information with selected contacts or public professional or personal purpose. In the first case, OSN is used as a facility oriented towards career management or business goals, so it is chosen to have a more serious image SNS, such as XING or LinkedIn because in this case, members know that the profession is influenced by OSN, they usually pay attention to the data content they publish about themselves and others. Introduction private use, they share more personal information, such as Contact data, personal photos or videos can be tagged. Share pictures with other members and automatically create links to their respective profiles [9], [10]. The core application used by OSN members is to create and maintain their contact lists, describe members' environments, and map them into a digital OSN graph by notifying SNS which Automatically updates based on contact profile changes, helps users stay up-to-date on their contacts, and a user's popularity is often measured by the number of contacts their profile is linked to. Analyzing the security of the OSN, the property, and the privacy of the users, some obvious threats become apparent. Often, a large amount of personal data of the participants on the website is stored at the provider, especially in the case of an OSN for non-professional purposes. This data is not available to the public. Visible, or, if the user is aware of privacy issues and able to use the settings of the corresponding SNS, to some extent select other members of the group. Since the proles are attributed to people who may 7 be known in the real world, they are implicitly assigned the same trust as the putative owner of the profile. In addition, any actions and interactions related to the profile are also attributable again to the putative owner of that profile, and different research shows that actors represent a security weak link in OSNs and that they are vulnerable to several types of social engineering attacks [12].

Online social networks (OSNs) have seen an incredible rise in popularity over the past several years because of rapidly developing technology. The ability of OSNs to provide a platform for people to communicate with their family, friends, and co-workers is the main reason why this phenomenon exists. Because social media and the media are so easily accessible, attackers are more likely to find and exploit information that is published on these platforms. There are a variety of ways that OSNs could be hiding their secrets and certainty from the public. When a user shares information online, it raises several security and privacy concerns [15, [17]. For example, if a user uploads images, videos, or audio which are then shared with others, this can lead to privacy concerns because people can access and view these materials without the user's knowledge. Additionally, sharing private information like this can also lead to security concerns because people who know about the information can potentially use it to harm or exploit the user. The attacker can use information shared by others for illegitimate purposes. If children are targeted, they are at even greater risk of being harmed or kidnapped [11], [14], [19]. This paper provides a comprehensive review of social network security threats and existing solutions that can provide security for social network users. We have discussed how various OSN web applications can be attacked by attackers by citing some statistical reports.

They have also discussed various defensive measures against OSN security threats. This paper discusses how to achieve trustworthiness in online social networks. People often take for granted the importance of protecting the information they keep on social networking sites because they view them as personal communication tools. People are increasingly posting information on social networks in a variety of formats, which could result in unprecedented access to personal and corporate data. Social network users retain a lot of information, making it particularly alluring for enemies looking to cause harm [13] [16]. With this large amount of knowledge at their disposal, they can wreak havoc on the world. 8 Social media is a great platform for marketers to advertise, but if they don't take social media security issues seriously, they leave themselves open to several dangers. Social networks can be divided into numerous varieties depending on their purposes. The four basic categories of social networks are "social relationships," "multimedia sharing," "professional," and "discussion forums." This section covers social networking sites that can be used to communicate with people, their vulnerabilities and incidents of phishing that have taken place on those sites. The website also includes a list of current security issues, as well as malicious content-based phishing attempts. According to the data shown below, social networking site use is growing rapidly, resulting in a large amount of data and information available on these sites [18], [20]. This has made it more difficult for people to keep track of important information and has created opportunities for cybercrime such as data interception, privacy spying, and information fraud. Some social media sites, such as Twitter, are not the best places to share personal information. However, some skilled attackers can still find out personal information by examining user posts and data they provide online. The private information we share online can be used by thieves to gain access to our email and passwords. We have restricted the study to a smaller number of networks to keep it.

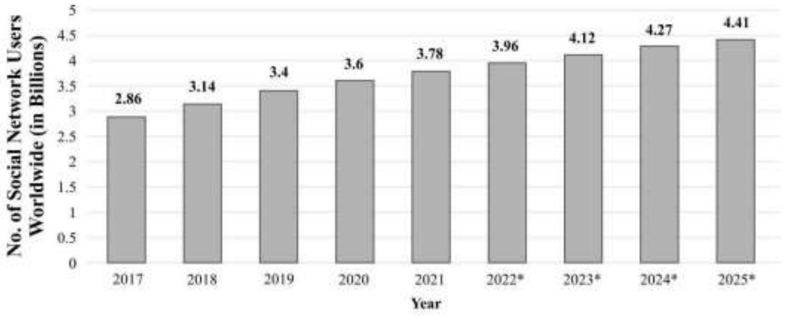

Fig 1. Number of Users on Social Media Worldwide (Year-wise)

As per Figure 1, the Social Networks (SNs) have attracted millions of users worldwide, 9 and they have evolved into an essential component of everyone's life in the modern world. There are many ways to define a social network. A social network is nothing more than a network of individuals, Webster claims. However, social networks are now making it very clear that they serve purposes beyond simply bringing people together. Networks can be referred to as online social networks because they are frequently used for online communication (OSN). No longer just a means of communication between people, the social networking sector has expanded, bringing its products and services into every aspect of our daily lives. Every day, new social networks evolve to meet diverse needs, each with a successful economic strategy. Friendster, Six Degrees, Orkut and other leading social networking platforms are just a few from India and Brazil, Orkut has emerged to the fore among OSN sites in 2007, in the year 1999, QQ - another popular instant messaging software - is launched. It has been widely used in China, however, Facebook and Instagram, which together have over 290 million users dominate the market today [23].

Online Social Networking (OSN) are becoming more and more popular among individuals all over the world Achieving data confidentiality from the user's perspective is of paramount importance due to the growing popularity. Although information such as photos or videos posted by the owner of a profile is visible to all friends, the user's preference for this information does not necessarily have to be viewed by everyone in the friend list A trust-based access control system is proposed to ensure data confidentiality. By restricting data access to a small group of selected friends. The degree of trust depends on a variety of factors [21], [22]. The study introduces the idea of trust - based on access control and suggests a way to deal with the threat to data confidentiality through friend sharing. The study found that sharing data in OSN can lead to data confidentiality attacks. It believe that a trust-based solution to the problem of data dissemination in the OSN using a trust-based method is the best way to solve the problem [27]. When it comes to Threat management, it is important to have a good understanding of how to protect one's data confidentiality and trust. Additionally, it is also important to be aware of the various online social networks that can be a threat to one's data privacy.

In section I introduce the social networking sites that can used to communicate with people, their vulnerabilities and incidence of phishing that have taken place on those sites. Section II contain the related work about the social network security. The section III contain proposed methodology of theoretical model, system model and algorithms. The section IV contain the experimental results and analysis of the results.

II. RELATED WORK

The concept of trust is multifaceted and context-specific [23]. Gefen et al. (2003) suggest that it is crucial to distinguish between trust as a belief in the positive attributes of the other party and trust as an intention to assume risk and make oneself 17 vulnerable to others when speaking to researchers of information systems (Mayer et al., 1995; Chopra and Wallace, 2003) [21], [23]. According to Dinev and Hart (2006) [27], it define trust in our study as the belief of the truster that the other party possesses qualities that prevent it from acting opportunistically (McKnight et al., 2002a; 2002b) [25].

It distinguish between three different categories because we are aware that there are many ways to categorize trusting beliefs: competence (the trustee's capacity to perform the tasks required by the truster), benevolence (the trustee's concern for and acting in the truster's best interest), and integrity (the trustee's honesty and commitment-keeping) (McKnight et al., 2002a) [24].

Studies focusing on e-commerce have found that trusting beliefs can have a favorable influence on the willingness to engage in the transaction by reducing the perceived risk's magnitude (e.g. Jarvenpaa and Tractinsky, 1999) [28]. The Social Exchange hypothesis suggests that trust can be used to lower the perceived costs of social transactions and motivate users to engage in them (Metzger, 2004). Gefen et al. (2003) mention that, in circumstances where risk is inherent to an activity, trust will serve as a risk-reducing strategy; risk, in turn, will directly impact behavior, even though the literature does not have a consensus answer on the relationship between trust, risk, and resulting behavior [30], [32].

Even while FG participants acknowledged that they were exposed to the OSN provider, they asserted that they had enough faith in Facebook to ensure that their data was handled responsibly. Their calculative notion that "if it gets out that they are utilizing our information, people will start to migrate to other networks" was a foundation for their trusting beliefs (FG quotation).

According to McKnight et al. (2002a), when users opt to provide the service provider access to their personal information, they will be more concerned with its goodness and integrity than with its expertise [33]. In our study, it only focus on beliefs about the OSN provider's goodness and integrity and exclude the evaluation of the OSN provider's competency from our conception of trust. It contend that consumers may feel less danger when disclosing personal information on a platform if an OSN provider is seen as being compassionate, honest, and consistent in its dealings with users [36].

A. *Revealing Hidden Attribute Values of Social Profile*

It's critical to have complete knowledge of a user's profile on an online social network to detect profile cloning. The mechanisms offered by online social networks provide us the option to show attribute values exclusively for user-defined groups of people in the network, which is the most common cause of hidden attribute values. Access Control Lists typically utilize this strategy (ACL). Despite their usefulness in safeguarding privacy, this circumstance makes our study more difficult in terms of detecting profile copying. To give analysis, it must first create a network with nodes that represent people and edges that define relationships between those people. Additionally, a set of functions on nodes that reflect social profile features and one function on edges that specifies relational force must be established. Here is selection of a node in that network and give it the designation by which it want to display concealed attribute data.

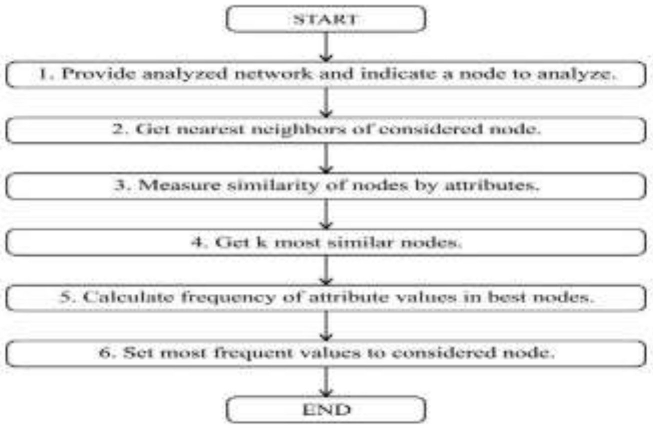
START
1. Provide analyzed network and indicate a node to analyze.
2. Get nearest neighbors of considered node.
3. Measure similarity of nodes by attributes.
4. Get k most similar nodes.
5. Calculate frequency of attribute values in best nodes.
6. Set most frequent values to considered node.
END

Fig 2. Online social network to detect profile cloning

B. *Node Similarity*

Following the disclosure of all attribute values in a user's social profile, it must offer a measure that enables us to identify the proper person from the pattern network in our analysed network. To accomplish this, it must employ a function that meets the requirements below: It should be simple to understand, allowing us to determine which personal profile, in terms of attribute values, is most similar to the profile of the person being considered without the need for any additional transformation of measuring function values. It should use a characteristic that is simple to obtain from the network model.

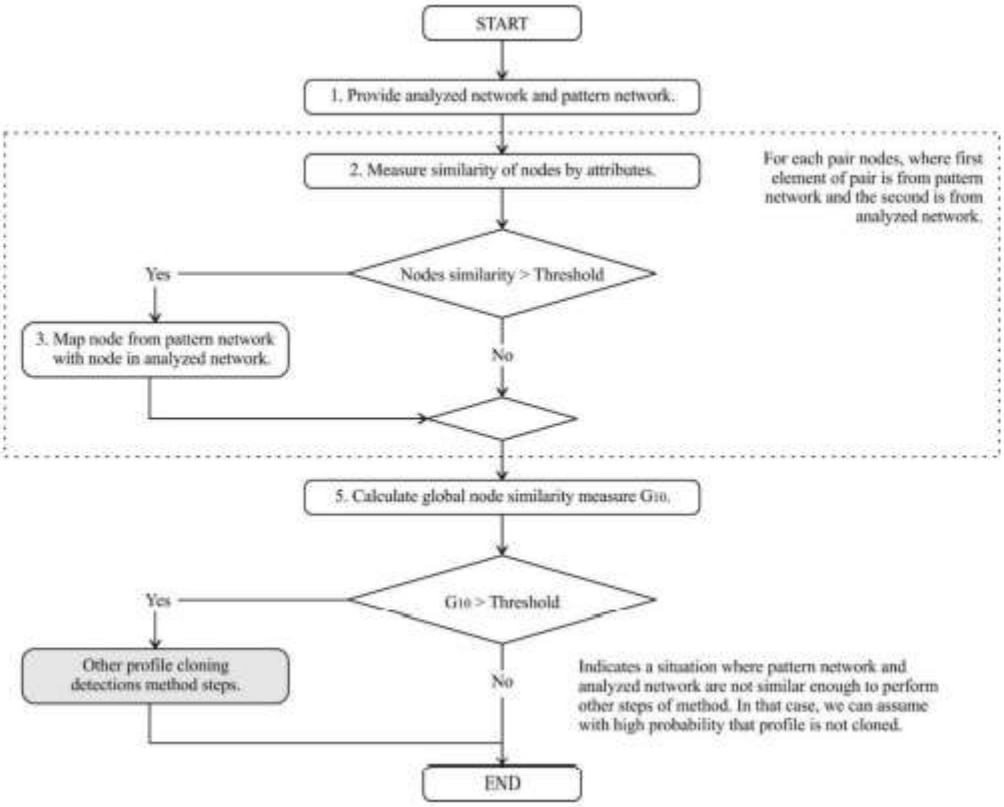
START
1. Provide analyzed network and pattern network.
2. Measure similarity of nodes by attributes.
For each pair nodes, where first element of pair is from pattern network and the second is from analyzed network.
Nodes similarity > Threshold
Yes
3. Map node from pattern network with node in analyzed network.
No
5. Calculate global node similarity measure G10.
G10 > Threshold
Yes
Other profile cloning detections method steps.
No
Indicates a situation where pattern network and analyzed network are not similar enough to perform other steps of method. In that case, we can assume with high probability that profile is not cloned.
END

Fig 3. Node Similarity

C. *Privacy-preserving social network analysis*

WBSN data is a valuable resource for social and marketing analysis, which can reveal information about how a social group has developed, how people work together to solve problems, how information is disseminated, and other topics so forth. They can also be used to improve social networking services and modify them following user preferences and interests. To the greatest extent feasible, it is vital to keep members' personal information private when evaluating WBSN data for statistical purposes.

This problem has so far been solved by anonym zing the network graph using one of two 30 primary techniques, namely edge perturbation or node anonymization. The former technique, known as naive anonymization, seeks to conceal the identities of members by assigning random identifiers to the appropriate network nodes. The idea of using approaches based on k-anonymity has been considered in cases where nodes are associated with qualities that can be used to identify the corresponding user. The use of the graph for network analysis is maintained while edge perturbation prevents an attacker from deducing the identities of network nodes based on the current associations by performing a set of random edge deletions and insertions.

It has been noted that the node anonymization options put out do not completely guarantee privacy protection. In particular, Backstrom, Dwork, and Kleinberg (2007) [30], conducted a thorough examination of the potential threats and asserted that interactive privacy protection measures are the most successful ones. In this method, the anonym zed network graph is not exposed; rather, it is examined by the social network management system after a question is submitted, and the outcome is then affected by introducing noise to the true answer.

When network data is studied using data mining tools, the objective of the available privacy-preserving strategies, both those based on graph anonymization and those limiting WBSN members' look ahead, is to protect users' privacy. Enabling a WBSN user to specify which information should be private or public and which members are authorized to access it is another issue. Current WBSNs apply very basic default protection mechanisms in this regard that cannot be modified by WBSN members.

D. *Privacy-Aware Access Control*

Resources in WBSNs must be protected, however, the current, imprecise implementations of WBSNs cannot fulfill this need. Therefore, an access control model for WBSNs considers the distinctive characteristics of the application domain to develop optimal access control approaches. In the sections that follow, the main requirements for an access control mechanism created especially for WBSNs are discussed. The current solutions are then reviewed.

According to the conventional approach, authorizations—which, in their most basic form, are tuples of the type’s s, p, and o—define access control criteria. S is the subject allowed access to object o under privilege p. A member might be required to update the authorizations governing his or her resources if they meet new people or if their participation in relationships ends, making such a system inappropriate for dynamic and distributed environments like WBSNs. By specifying the requirements users must meet to access a certain resource, it is preferable in this circumstance to expressly identify authorized users.

Several access control models have been put forth thus far that indicate authorized users by their attributes rather than only by their identities. The most common model is the one based on credentials. Certificates (For WBSNs, an analogous strategy can be used.

In actuality, WBSN members frequently create materials with a particular audience in mind, such as their friends or coworkers. Relationships can therefore be utilized to specifically identify authorized members in a WBSN environment.

Addressing two key problems is necessary for relationship-based access control enforcement. To prevent security attacks based on fabricating relationships, it is first necessary to be able to confirm the dependability and validity of relationship-related information. Second, privacy protection laws may apply to relationship information; as a result, methods to control disclosure should be in place.

Both Carminati et al. and Hart et al. [32], [34], techniques presuppose that interactions are open to the public. The earlier study by suggesting a privacy-aware access control mechanism in which the existing relationships are safeguarded by a set of rules known as distribution rules. The distribution of relationship certificates to approved members is governed by such regulations. When enforcing access control, discuss the issue of safeguarding relationship information that might be implied by access restrictions. More specifically, each WBSN participant m owns a key for the particular kind of 32 relationship in which they engage. All WBSN members in or her social network group, or all WBSN members connected to m via pathways identified with those connection categories, share these keys. When m receives a request for access to a resource that belongs to him or her, he or she does not communicate the associated access rules in plaintext. Instead, the corresponding relationship key is used to encrypt each access condition in the access rule. For instance, m will encrypt it with the key associated with that relationship type if an access condition restricts relationships of type of Friend.

III. PROPOSED METHODOLOGY

Users can access social networking services from social network providers, who may also offer additional interfaces and services to other clients These clients could represent distinct industries and have different objectives Sponsors in particular are clients who utilize the OSN platform to sell their services to consumers Their advertising can take many different forms Plain commercial sponsors can advertise their products by purchasing banner space or other marketing services from the SNP SNS frequently have "market pages" where users can post classified ads (ads), job offers, and the like for which they may be charged In the OSN, sponsors may also create commercial interest groups or profiles.

Third-party service providers are another kind of OSN users who add their programs to increase the capabilities and content of SNS Quizzes and games are examples of apps that are often run on servers controlled by third parties that are connected to the SNS via the proper APIs These programs frequently have significant access to OSN users' personal information.

To partially conceal user’s presence in the OSN, an OSN user can set output limits to this function. The OSN's profile browsing feature would still allow access to the protected profile, though. Nevertheless, by placing limitations on the output of the contact list browsing feature, sensitive relationships can be concealed from unauthorized users. This constraint can also totally hide some profiles from the OSN when combined with the limitations on profile lookup, as these profiles will no longer be accessible to users outside of their contact list. Note that new contacts may still be added to the contact list of the profile owner at their discretion. Another illustration is the ability of the profile owner to control whether or not their profile is disclosed to other users through the control over the output of the profile retrieval function. This enables some OSN users to conceal some 131 private profile details from certain partners. Finally, restrictions on a wide range of networking and data functions can be used to secure the data associated to online or online indicators, one-to-one or one-to-many communications, such as postings, walls, comments, positive or negative markings, tags.

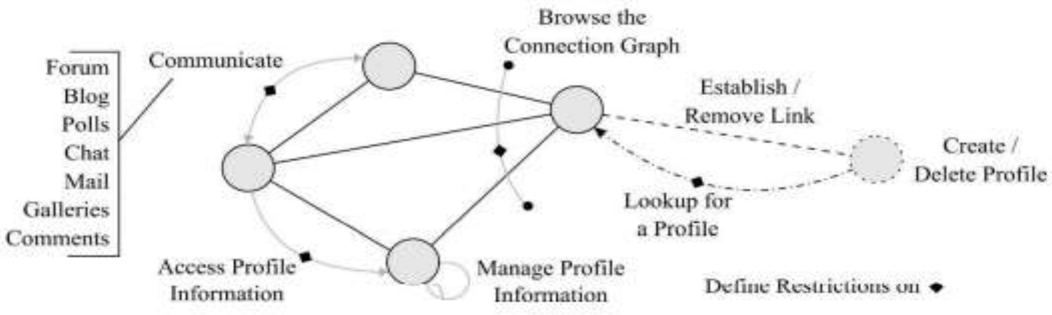

Fig 4. OSN Network

A. *Approach To Privacy As Protection Module*

The goal of PETs (“Privacy Enhancing Technologies”) in the context of OSNs is to enable individuals to engage with others, share, access, and publish information online, free from surveillance and interference Ideally, the only information that a user explicitly shares is available to her intended recipients, while the disclosure of any other information to any other parties is prevented Furthermore, PETs aim to enhance the ability of a user to publish and access information on OSNs by providing her with means to circumvent censorship.

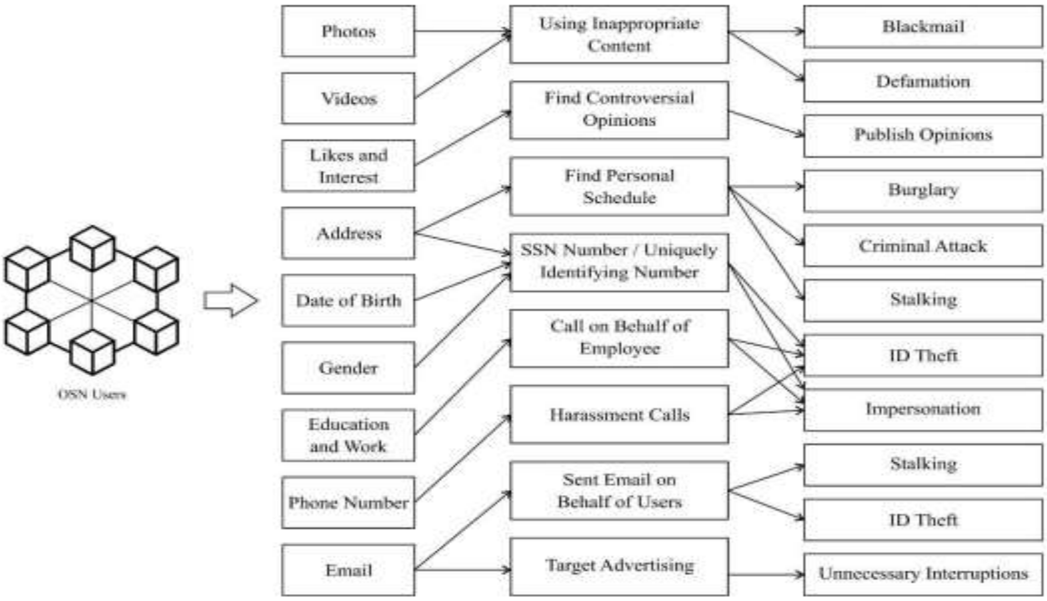

Fig 5. Information on OSN Users

B. *Structural model evaluation*

Using SmartPLS's bootstrapping mode, the path validity coefficient in the model is tested in this part. Which pathways are significant is determined by looking at the P-value. The specifics that show which relationships are important in the model are shown in Figure 4.10.

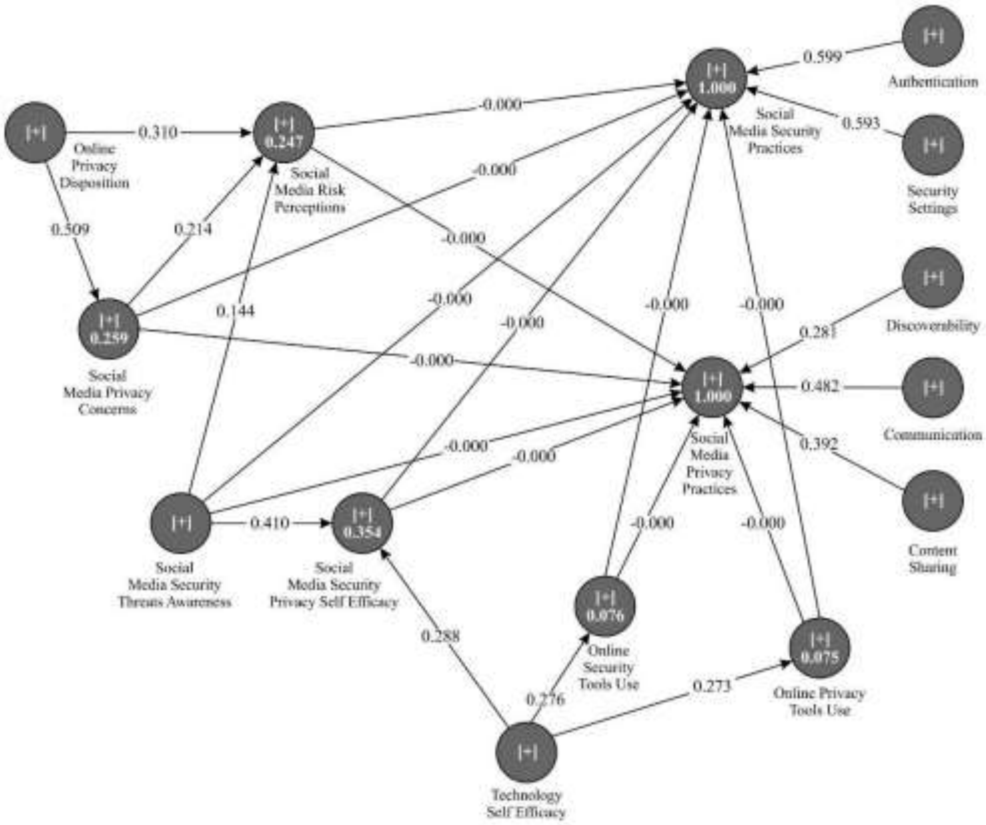
0.310
0.247
Online Privacy Disposition
Social Media Risk Perceptions
0.509
0.214
0.144
0.259
Social Media Privacy Concerns
-0.000
1.000
Social Media Security Practices
0.599
Authentication
0.593
Security Settings
0.281
Discoverability
0.482
Communication
0.392
Content Sharing
1.000
Social Media Privacy Practices
0.410
0.354
Social Media Security Threats Awareness
Social Media Security Privacy Self Efficacy
0.288
0.076
Online Security Tools Use
0.276
0.273
0.075
Online Privacy Tools Use
Technology Self Efficacy

Fig 6. Structural Model Result

C. - *oigh algorithm*

We note that the generalization lattice's unique properties for IGH data are as follows.

1. A k-anonymous node is defined as all direct generalization nodes in the higher generalization lattice level.
2. A non-anonymous node is defined as all direct generalization nodes in the lower generalization lattice level.
3. The nodes at the same generalization lattice level have the same precision.
4. A higher generalization lattice level always results in a node with more precision than a lower generalization lattice level.

The best k-anonymous node is always found among the nodes in the lowest k-anonymous level, or the level of the lattice where the k-anonymous nodes are located, as a result of all these unique qualities. We suggested Extended-OIGH technique is based on a tree search, and it finds the kanonymous node with the lowest precision by conducting an in-order depth-first search. According to the unique properties of an IGH data's optimal k-anonymity, the best answer is always found at the lowest level of the k-anonymous node. We take advantage of these qualities when designing the algorithm

```
Input: lowest level found k-anonymous node L

Output: optimal k-anonymity node op

begin
    $R = Traversal\ Route(L)$
    if !R then
        $OP \leftarrow$ an optimal node among the k-anonymous nodes in level L
        Return OP

    else
        for each node in R do
            if node is not tagged then
                if node is k-anonymous node
                    Mark node as k-anonymous node
                    Tag all successor nodes as k-anonymous
                    L ← node level
                    Exit
                else
                    Mark node as non-anonymous node
                    Tag all predecessor nodes as non-anonymous
                End
            end
        End
        Extended – OIGH(L)
    End
End
```

The Extended-OIGH method, as indicated in Algorithm 1, is where the algorithm first starts, taking as an input the lowest level discovered k-anonymous node L. The input L will 226 first be established as the top level of the generalization lattice. When this happens, Algorithm 2's Traverse Route sub-algorithm is called to provide a route from the root node 000> at level 0 to a node at level L. The technique iteratively calculates each node's k- anonymity from the path, starting from the node at the highest level.

All direct generalization nodes in the upper level are designated as k-anonymous nodes and level L is set as a k-anonymous level if the node is a k-anonymous node. All direct generalization nodes at the lower level are labeled as non-anonymous nodes if the node is not a k-anonymous node. Using new input L, the algorithm will keep doing the Extended-OIGH until all nodes above and below the k-anonymous level have been tagged. The algorithm will only evaluate the nodes at and below the k-anonymous level since the optimal node is always at the lowest level detected k-anonymous nodes. As a result, it might skip certain nodes above the k-anonymous level.

The Traverse Route algorithm, which searches the boundary of the generalized lattice between the higher and lower levels alternately, achieves an in-order traversal. As a result, the number of unexplored nodes in the lattice may be more than that of unexplored nodes found using pre-order or post-order traversals, which may waste execution time by traversing up or down until the unexplored node is first discovered.

D. Example of Extended-OIGH Algorithm

To demonstrate our suggested work, we provide an example. Assume that we wish to publish a 2-anonymous IGH dataset that is optimal. The dark nodes with the bold contour on the lattice of generalization are the k-anonymous nodes. Starting with node "000," the Extended-OIGH algorithm evaluates the network. Next, nodes 222, 221, and 220 are handled using the in-order traversal methods.

These two nodes are labeled because they are the k-anonymous nodes 222 and 221. The current k-anonymous level is 5. Since node 220 is a non-anonymous node, nodes 210, 200, 110, 200, 100, and 010 which are part of direct generalization are also designated as non-anonymous nodes.

Algorithm 1: THE fednoise ALGORITHM

```
Input: the number of clients n
Output: trainSet; testSet; evaluation result; training timefor $t = 1$ to T do
    Server broadcasts $w_i$;
    for client $i \in [n]$ do
        $w_{i,0}^{i} \leftarrow w_i$;
        for $t = 0$ to $K - 1$ do
            $w_{i,\tau+1}^{i} = w_{i,\tau}^{i} - \xi \nabla f_i \left( w_{i,\tau}^{i} \right)$
        end for
        $w_{\tau+1}^{i} = w_{\tau}^{i} + N_i$
        Sends $w_{\tau+1}^{K}$ to server;
    end for
end for
```

Algorithm 2: split & carry algorithm

The Split & Carry method is the first useful algorithm based on the initial workable solution. Rows in a sorted matrix are probably equivalent rows in an optimally k- 230 anonymized dataset, therefore we use the initial solution to divide the main problem into smaller sub-problems with manageable sizes. As a result, we predict that rows in a sorted matrix that are far apart are unlikely to be equivalent and can be divided into various sub problems that the general model can handle. Our Split & Carry algorithm, which is detailed in Algorithm 1, is based on this concept. The algorithm's input values are as follows: 1. arrT an array describing the data type of each column (from {Integer, Continuous}, used in Gurobi solver)

```
Data: $x, U, L, W, k, S, arrType$
Result: A, f
$VAR = [VAR]_{j \in J}$ ← compute variances of all attributes;
$\bar{x} \leftarrow sort(x, VAR)$;
$C_0 \leftarrow \underline{0}$ ;
f ← 0 ;
for $m := 1$ to $\left\lceil \frac{n}{\lfloor k \times S \rfloor} \right\rceil$ do
    $Sub_m \leftarrow C_{m-1}$ + Read the next $k \times S$ rows from $\bar{x}$;
    $(A_m, f_m) \leftarrow Solve(Sub_m, arrType)$ optimally
    $f \leftarrow update(f, f_m)$ ;
    $C_m \leftarrow$ rows in equivalence classes of last k rows in $A_m$;
    $A \leftarrow update(A, A_m)$
    $m = m + 1$
end
return $(A, f)$
```

Here is employ the parameter S to change the start size of the sub-problems in addition to the first five input values, which are shared by the optimal model. S captures the smallest number of k-sets that must be contained in each sub-problem. Because there is only one simple k-set when S = 1, S is a user parameter with a value greater than 2. The chosen value for S is determined by the process's utility and intended efficiency. Since S specifies the minimum size of the sub-problem, a big value will raise the difficulty of each sub-problem.

All direct generalization nodes in the upper level are designated as k-anonymous nodes and level L is set as a k-anonymous level if the node is a k-anonymous node. All direct generalization nodes at the lower level are labeled as non-anonymous nodes if the node is not a k-anonymous node. Using new input L, the algorithm will keep doing the Extended-OIGH until all nodes above and below the k-anonymous level have been tagged. The algorithm will only evaluate the nodes at and below the k-anonymous level since the optimal node is always at the lowest level detected k-anonymous nodes. As a result, it might skip certain nodes above the k-anonymous level.

When S is little, we are restricting the potential space from which we can build k-sets. In our studies, we will show that, when k is small and the number of records is big, a modest S is sufficient. In general, because MILP solvers are frequently multi-threaded, the running times of the sub-problems also depend on the availability of computational resources. When we have a large dataset, we are likely to find many records that are similar, which means each sub-problem to be solved as a small solution space.

I. Experimental result

The 630 people in the final dataset were classified using the following criteria: 60% of the population is male, 40% is female, 40% has a graduate degree, 52% has an undergraduate 155 degree, 8% has completed their secondary education, and the age ranges from 18 to 65 with the 26 to 35 age group being the largest. In addition, we made an effort to find participants from various areas so that our study wouldn't be centered on a particular cultural background. We study security and privacy behaviors based on a global trend thanks to the diversity of answers from a geographic standpoint, which enables us to look into the likelihood of variations in user behavior in various contexts. Male respondents performed better on average than female respondents in all areas (Technological Self-Efficacy) and social network characteristics (Figures 7 and 8), as shown (Social Media Security & Privacy Self-efficacy, and Awareness). A similar pattern happens in two variables of Security practices, suggesting that males have a greater rate of self-claimed security practice.

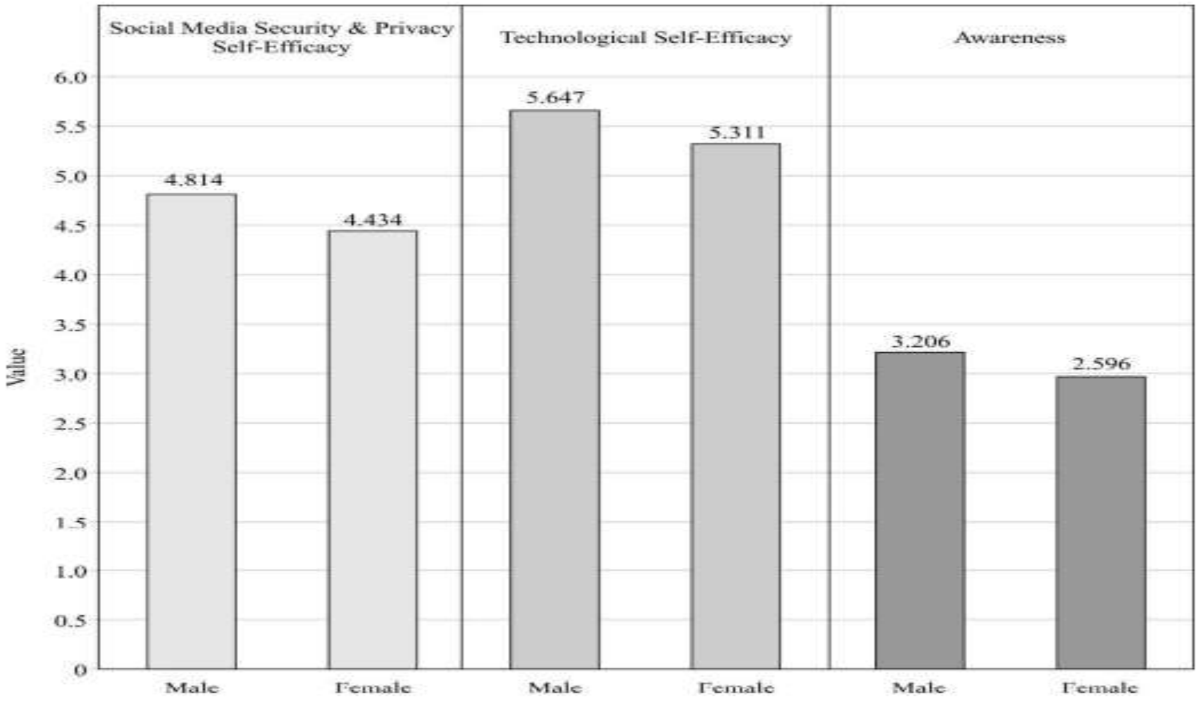

Fig 7. Average Rate of Major Proficiency Variables Based on Gender

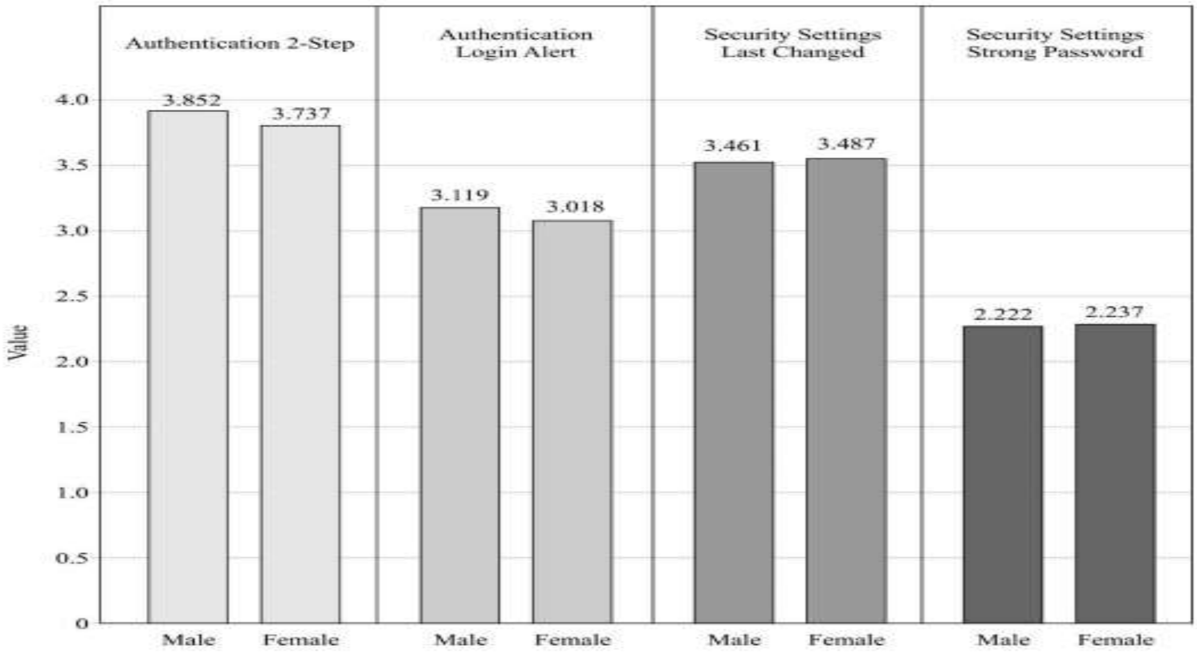

Fig 8. Average Rate of Security Practices Based on Gender

As seen in Figure 9, when females exhibit a higher proportion of behavioral traits, the tendency is the contrary concerning social media privacy practices. This suggests that when it comes to privacy-related online behavior, women are generally more circumspect. In addition to gender, the educational attainment of end users may also be a sign of prejudice. Figure 10 shows that the disposition, concern, and perception of risk increase with more educational attainment.

In addition to gender, the educational attainment of end users may also be a sign of prejudice. Figure 10 shows that the disposition, concern, and perception of risk increase with more educational attainment.

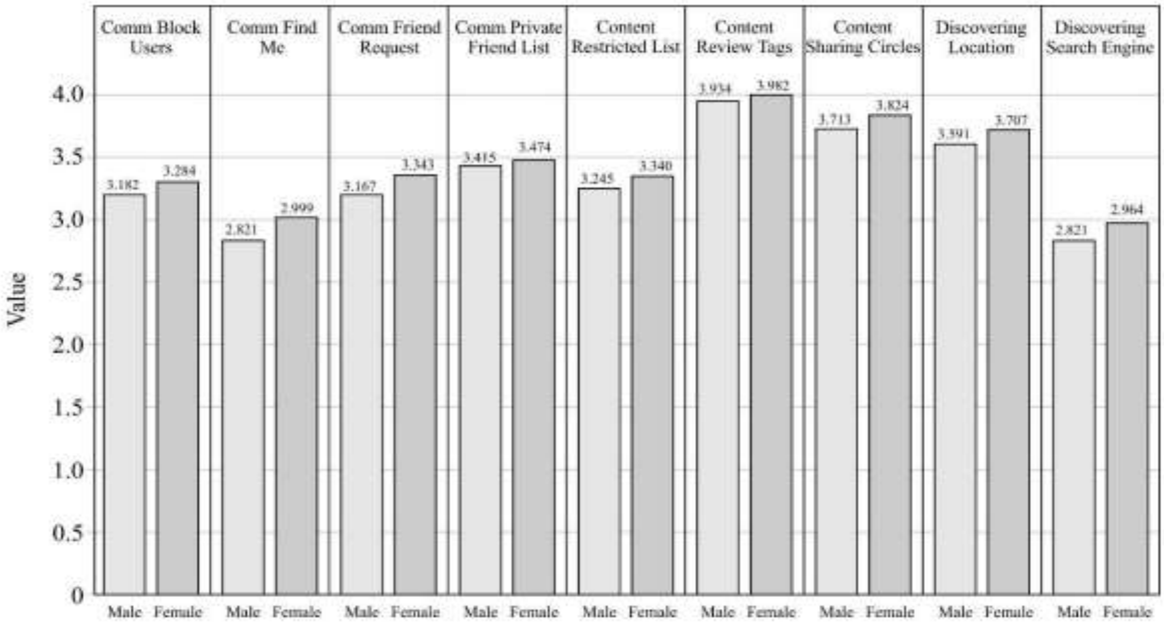

Fig 9. Average rate of Privacy Practices in Social Media based on Gender

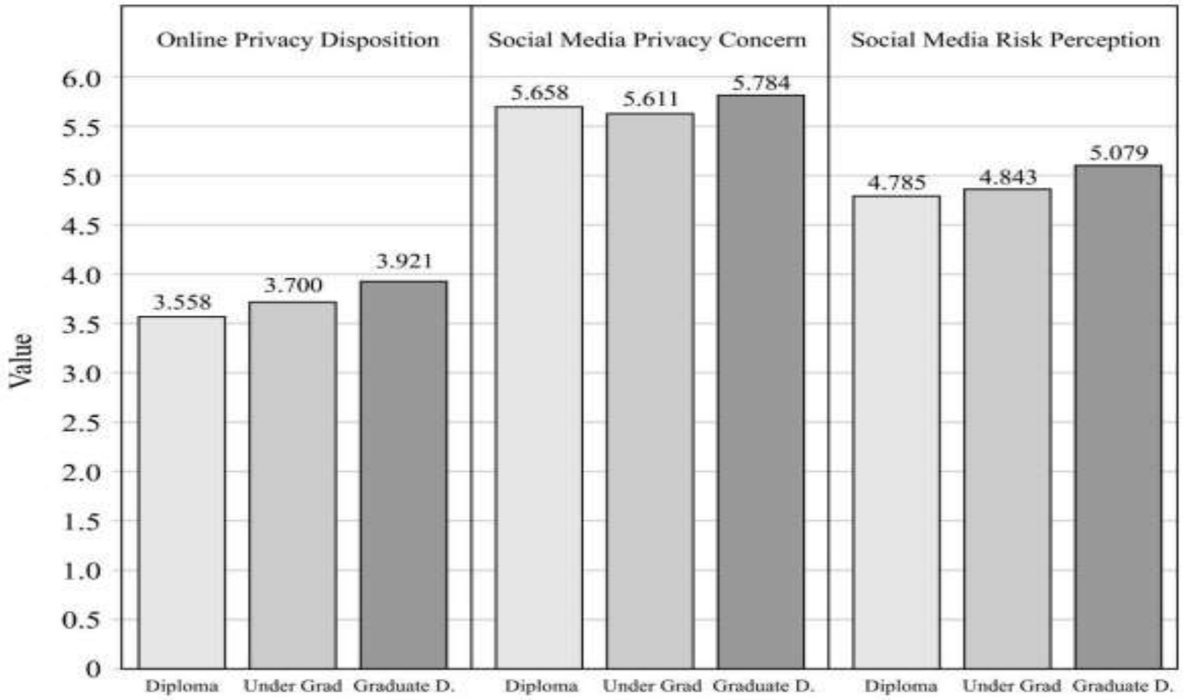

Fig 10. Average Rate of Posture Variables Based on Degree Status

Technologically speaking, most users often have two to three social networks (Figure 10). The most intriguing finding from this data is how many respondents—148 out of 630, or 22.2%—have five or more social media accounts. Additionally, end users typically devoted 1 to 10 hours per week to social media (Figure 11).

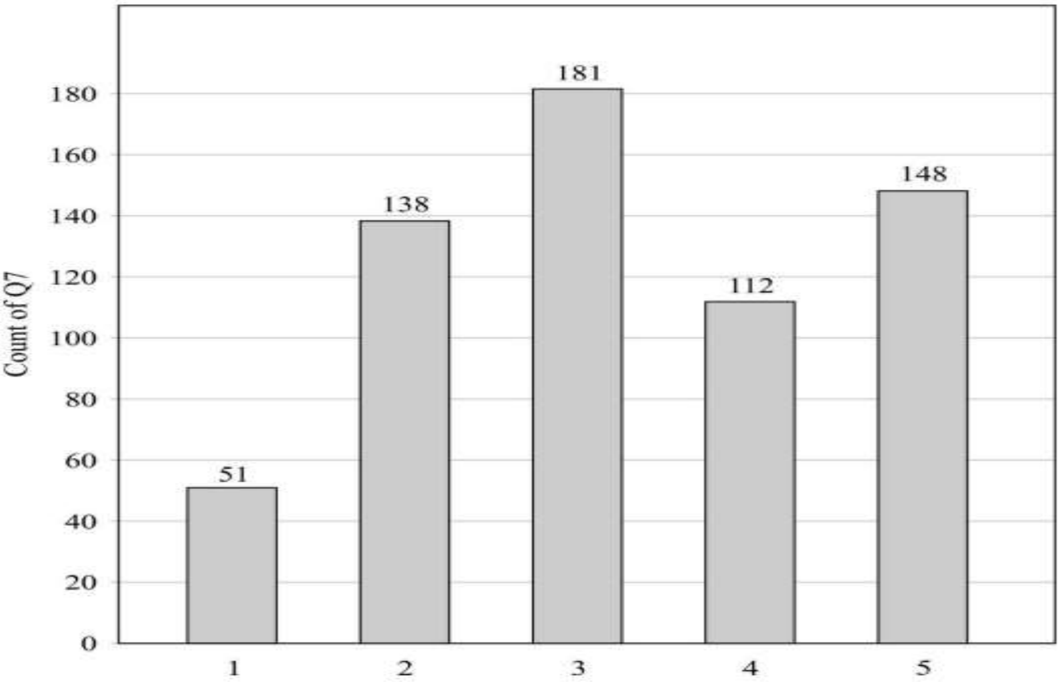

Fig 11. Number of Social Platforms Used by Users

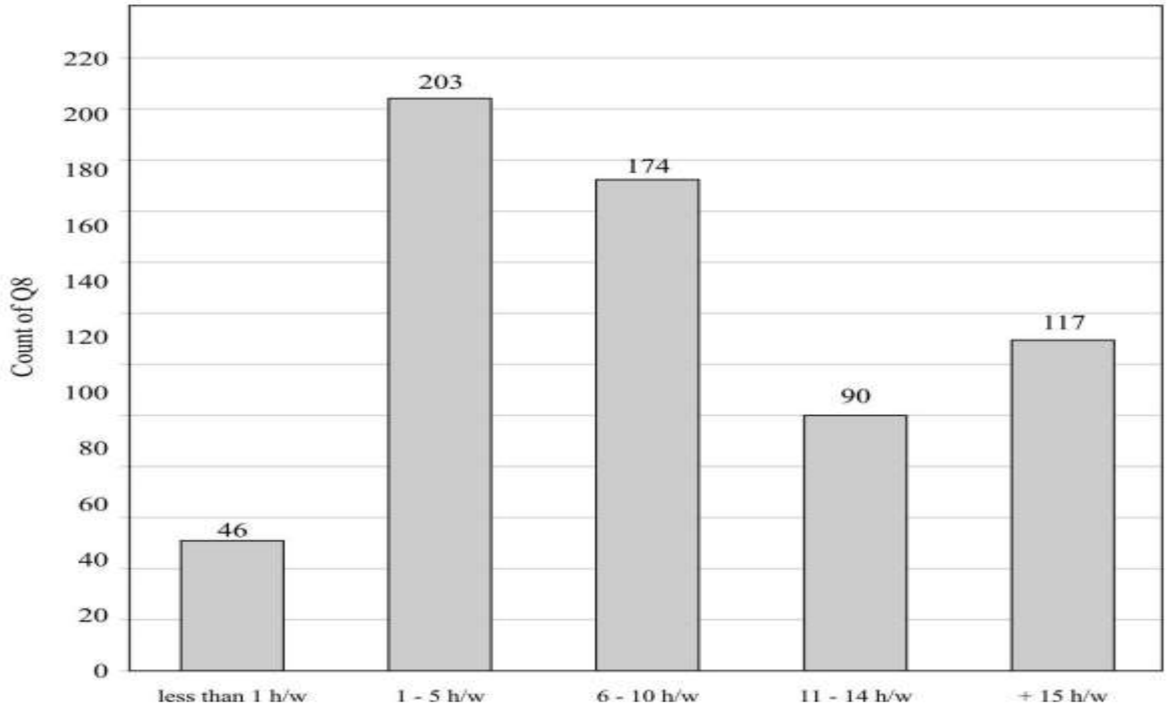

Fig 12. Total Hours Spent on Social Media per Week

End users typically have a very small or huge network size for their initial platform, as seen in Figure 12, which demonstrates the vastly varying capabilities they anticipate from various social networks. A varied range of end-user connections is seen in the fact that about 24% of end users have more than 500 connections on their first platform of choice. However, as shown in Figure 13, end users are reluctant to share excessive amounts of information due to the propensity for large-scale networks.

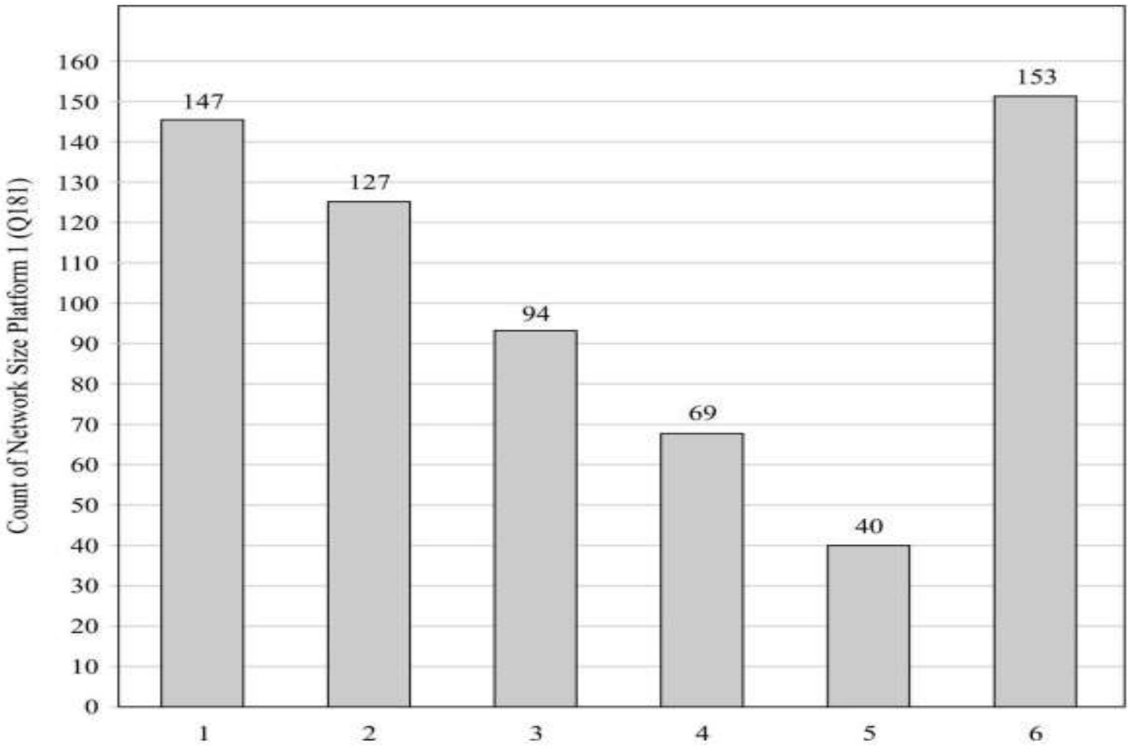

Fig 13. Network Size for First Platform

Measurement Model Evaluation

The reflective components in this study will be validated using methods including composite reliability, AVF, and Cronbach's alpha, while the formative constructs are 164 validated using collinearity evaluation and significance & relevance assessment for outer weights.

Measurement Model Assessment

Discriminant Validity at Construct level by analysing correlations between reflective variables, the next step is to analyse discriminant validity among constructs. The FornellLarcker criterion dictates that this be done by contrasting the estimated correlations with the square root of the average variance extracted (AVE). According to Table 1, the square roots of AVE are larger than the correlation of the same constructs with other constructs, indicating that the model's discriminant validity is acceptable. The construct explains more than half of its indications when the AVE value is 0.50 or above. Conversely, AVE values 165 of less than 0.50 denote that there is still more variance in the error than has been adequately explained.

TABLE I
AVERAGE VARIANCE EXTRACTED AND INTER-CONSTRUCT CORRELATIONS

Measurement Items	Model Reflective constructs							
	OPR	**OPTU**	**OSTU**	**SMPC**	**SMRP**	**SMSPSE**	**SMSTA**	**TSE**
Online Privacy Disposition	0.837	-	-	-	-	-	-	-
Online Privacy Tools Use	0.076	1	-	-	-	-	-	-
Online Security Tools Use	0.14	0.484	1	-	-	-	-	-
Social Media Privacy Concerns	0.509	0.138	0.164	0.829	-	-	-	-
Social Media risk Perceptions	0.436	0.181	0.179	0.386	0.832	-	-	-
Social Media Security privacy Self-Efficacy	0.121	0.363	0.407	0.064	0.092	0.854	-	-
Social Media Security Threats Awareness	0.124	0.382	0.437	0.104	0.204	0.535	0.855	-
Technology Self Efficacy	0.131	0.273	0.276	0.131	0.165	0.466	0.434	0.917

Measurement Model

Convergent Validity Cronbach's alpha, composite reliability, and AVE are the three metrics used to evaluate convergent validity. The internal consistency reliability will be assessed in the first stage. Based on table 2 the interconnectedness of the indicators, Cronbach's alpha is used to assess the dependability of a group of construct indicators. Several 0.70 or higher is regarded as a reliable indicator of internal consistency. Additionally, the overall model build dependability is greater than 0.7. Assessing AVE is the last stage in analyzing convergent validity. All of the constructs have rates over 0.68 (nearly all of them above 0.7), which guarantees their dependability and reflection in the model.

TABLE II
CONSTRUCTS STATISTICS CONVERGENT VALIDITY

Construct	**Cronbach's Alpha**	**Composite Reliability**	**Average variance Extracted (AVE)**
Online Privacy Disposition	0.788	0.875	0.7
Online Privacy Tools Use	1	1	1
Online Security Tools Use	1	1	1
Social Media Privacy Concerns	0.886	0.916	0.686
Social Media risk Perceptions	0.889	0.918	0.692
Social Media Security privacy Self-Efficacy	0.907	0.931	0.729
Social Media Security Threats Awareness	0.908	0.932	0.732
Technology Self Efficacy	0.905	0.94	0.84

II. CONCLUSIONS AND FUTURE SCOPE

This paper provide the comprehensive review of social network security threats and existing solution that can provide the security for the social network user. The proposed model provide the functionality like, revealing hidden attribute value of social profile, node similarity, privacy- preserving social network analysis and privacy-aware access control. In this paper the mainly uses the approach to privacy as a protection model, structural model evaluation, OIGH algorithm. When a match between users is found, further protocols should facilitate additional functionalities, such as gradual information disclosure, or exchanging recommendations in a privacy-friendly way. The paper including all the major proposed algorithms for data confidentiality and security in online social networks. It focuses on various algorithms which can help identify the threat and give more security to users without any information breach.

REFERENCES

[1] Vasilyevna, N.B.; Sang-Soo Yeo; Eun-Suk Cho; Jeon-Ah Kim, "Malware and Antivirus Deployment for Enterprise IT Security", in IEEE, Ubiquitous Multimedia Computing, 2008. UMC '08. International Symposium. Page(s): 252-255.

[2] Marpaung, J.A.P.; Sain, M.; Hoon-Jae Lee, "Survey on malware evasion techniques:
State of the art and challenges" in IEEE, Advanced Communication Technology (ICACT), 2012 14th International Conference. Page(s): 744-749.

[3] Gamayunov, D., "Towards Malware-Resistant Networking Environment", in IEEE, SysSec Workshop (SysSec), 2011 First. Page(s): 79-82.

[4] Hashimoto, G.T., "A Security Framework to Protect Against Social Networks Services Threats", in IEEE, Systems and Networks Communications (ICSNC), 2010 Fifth International Conference. Page(s): 189-194.

[5] https://techcrunch.com/story/facebook-responds-to-data-misuse/

[6] Fire, Michael, Roy Goldschmidt, and Yuval Elovici. "Online social networks: threats and solutions." IEEE Communications Surveys & Tutorials 16.4 (2014): 2019-2036.

[7] Kayes, Imrul, and Adriana lamnitchi. "A survey on privacy and security in online social networks." arXiv preprint arXiv: 1504.03342 (2015).

[8] Aljohani, Mashael, Alastair Nisbet, and Kelly Blincoe. "A survey of social media users privacy settings & information disclosure." (2016).

[9] Ajami, Racha, et al. "Security challenges and approaches in online social networks: A survey." IJCSNS 11.8 (2011): 1.

[10] Zheleva, Elena, and Lise Getoor. "Privacy in social networks: A survey." Social network data analytics. Springer, Boston, MA, 2011. 277-306.

[11] Cavoukian, Ann. "Privacy in the clouds." Identity in the Information Society 1.1 (2008): 89-108.

[12] Chewae, Mafaisu, et al. "How Much Privacy We Still Have on Social Network?" International Journal of Scientific and Research Publications 5.1 (2015): 2250-3153.

[13] Rubinstein, Ira. "Big data: the end of privacy or a new beginning?" (2012).

[14] Jayalakshmi, N., and R. G. Kavitha. "A Survey on Privacy in Social Networking

Websites." Adarsh Journal of Information Technology 5.2 (2016): 30-37.

[15] Zhang, Chi, et al. "Privacy and security for online social networks: challenges and opportunities." IEEE network 24.4 (2010).

[16] Kayes, Imrul, and Adriana lamnitchi. "Privacy and security in online social networks: A survey." Online Social Networks and Media 3 (2017): 1-21.

[17] Gao, Hongyu, et al. "Security issues in online social networks." IEEE Internet Computing 15.4 (2011): 56-63.

[18] Wang, Yang, and Alfred Kobsa. "Privacy in online social networking at workplace." Computational Science and Engineering, 2009. CSE'09. International Conference on. Vol. 4. IEEE, 2009.

[19] Toch, Eran, Yang Wang, and Lorrie Faith Cranor. "Personalization and privacy: a survey of privacy risks and remedies in personalization-based systems." User Modeling and User-Adapted Interaction 22.1-2 (2012): 203-220.

[20] Stutzman, Fred, Robert Capra, and Jamila Thompson. "Factors mediating disclosure in social network sites." Computers in Human Behavior 27.1 (2011): 590-598.

[21] Madejski, Michelle, Maritza Lupe Johnson, and Steven Michael Bellovin. "The failure of online social network privacy settings." (2011).

[22] Li, Li, Alexandre Bartel, Tegawendé F. Bissyandé, Jacques Klein, Yves Le Traon, Steven Arzt, Siegfried Rasthofer, Eric Bodden, Damien Octeau, and Patrick McDaniel. "Iccta: Detecting inter-component privacy leaks in android apps." In Proceedings of the 37th International Conference on Software Engineering-Volume 1, pp. 280-291. IEEE Press, 2015.

[23] Savla, P., & Martino, L. D. (2012). Content analysis of privacy policies for health social networks. Proceedings 2012 IEEE International Symposium on Policies for Distributed Systems and Networks, POLICY 2012, 94-101. http://doi.org/10.1109/POLICY.2012.20

[24] Alsagri, H. S., & Alaboodi, S. S. (2015). Privacy awareness of online social networking in Saudi Arabia. 2015 International Conference on Cyber Situational Awareness, Data Analytics and Assessment, CyberSA 2015. http://doi.org/10.1109/CyberSA.2015.7166111

[25] Khan, R., & Hasan, R. (2016). The Story of Naive Alice: Behavioral Analysis of Susceptible Internet Users. Proceedings International Computer Software and Applications Conference, 1, 390-395. http://doi.org/10.1109/COMPSAC.2016.206

[26] Kumaraguru, P. (2012). Privacy in India: Attitudes and Awareness V 2.0

[27] Dhawan, S., Singh, K., & Goel, S. (2014). Impact of privacy attitude, concern and awareness on use of online social networking. Proceedings of the 5th International Conference on Confluence 2014: The Next Generation Information Technology Summit, 14-17. http://doi.org/10.1109/CONFLUENCE.2014.6949226

[28] Zeadally, S., & Winkler, S. (2016). Privacy Policy Analysis of Popular Web Platforms. IEEE TECHNOLOGY AND SOCIETY MAGAZINE, (june), 75-85.

[29] Susan Farrell. (2016). 28 Tips for Creating Great Qualitative Surveys. Retrieved May 29, 2017, from https://www.nngroup.com/articles/qualitative-surveys/

[30] Yang, Yanjiang. Haibing Lu, and Jian Weng. "Multi-user private keyword search for cloud computing." Cloud Computing Technology and Science (CloudCom), 2011 IEEE Third International Conference on. IEEE, 2011.

[31] Zhibin Zhou and Dijiang Huang. Efficient and secure data storageopera- tions for mobile cloud computing. In Proceedings of the 8thinternational Conference on Network and Service Management, pages3745. Interna- tional Federation for Information Processing, 2012.

[32] Han, Jinguang, Willy Susilo, and Yi Mu. "Identity-based data storage in cloud computing." Future Generation Computer Systems 29.3 (2013): 673-681.

[33] Li, Jin, et al. "Identity-based Encryption with Outsourced Revocation in Cloud Computing" (2015). 7 Li, Jin, et al. "Identity-based Encryption with Outsourced Revocation in Cloud Computing" (2015).

[34] Hur, Junbeom, and Dong Kun Noh. "Attribute-based access control with efficient revocation in data outsourcing systems." IEEE Transactions on Parallel and Distributed Systems 22.7 (2011): 1214-1221.

[35] Mahalle Parikshit, et al. "Identity establishment and capability based access control (IECAC) scheme for Internet of Things." WPMC. 2012.

[36] Sahai, Amit, and Brent Waters. "Fuzzy identity-based encryption." Annual International Conference on the Theory and Applications of Cryptographic Techniques. Springer, Berlin, Heidelberg, 2005.

[37] Karati, Arijit, et al. "Provably secure and lightweight identity-based au- thenticated data sharing protocol for cyber-physical cloud environment." IEEE Transactions on Cloud Computing (2018)