Privacy Impact Assessment for Online Social Networks

Yong Wang
College of Business and Information Systems
Dakota State University
Madison, SD 57042
yong.wang@dsu.edu

Raj Kumar Nepali
College of Business and Information Systems
Dakota State University
Madison, SD 57042
rknepali@pluto.dsu.edu

***Abstract* – Many threats and attacks have been found in online social networks. When a security incident happens and data loss occurs, it is important to assess how the lost data affects user privacy. Most approaches for privacy impact assessment are based on checklists and auditions. There is lack of quantitative analysis approach to study privacy impact. Privacy impact assessment is a very challenging issue. First, data loss includes direct data loss, indirect data loss, and potential data loss. The impact of these data loss to user privacy should all be considered. Second, privacy impact assessment requires measuring privacy. Privacy measurement itself is a challenging issue. Third, users are all connected in a social network. Data loss may spread and propagate across the whole social network. In this paper, we summarize issues and challenges for privacy impact assessment. We further propose a quantitative analysis approach to assess privacy impact for online social networks. Two particular challenges are considered in the paper, privacy impact assessment when partial user information is disclosed, and privacy impact assessment when a group of user accounts are compromised. The paper provides a quantitative analysis approach for government agencies, enterprises, and organizations to assess privacy impact for online social networks when a security incident occurs.**

Keywords – online social networks, data loss, privacy impact assessment, privacy measurement

I. Introduction

Online social networks (OSNs) such as Facebook have attracted millions of users worldwide. According to Facebook, as of September 30, 2013, there were 1.19 billion monthly active users [1]. As more individuals and organizations share information in social networks, it also raises many security and privacy concerns. A social network is a place where people are willing to share information. The shared information is generally available to public and can be retrieved by users around the world. However, the shared information may also include users' personal data and could be manipulated by a malicious user against a person. Social networks are also widely used by organizations for business. In a survey on 250 of Fortune's Most Admired U.S. Companies, 91% of the organizations surveyed use at least one type of social media, e.g., Facebook, Twitter, or YouTube [2]. Social network security breaches have become one of the biggest threats to organizations and have resulted in millions of dollar in loss [3]–[5]. Associated Press's twitter account was hacked on April 23, 2013 and tweeted about explosion in the White House and resulted in $136 billion worth of fall in stock market [3]. In another survey commissioned by Check Point Software Technologies, 48 percent of enterprises surveyed claimed that they had been targeted with more than 25 successful social engineering attacks in the past two years [6]. For each social engineering attack, companies incurred losses between $25,000 and $100,000 [6].

Many threats and attacks have been found in online social networks. Examples of these threats and attacks include, but are not limited to, inference attacks, impersonation attacks, sybil attacks, phishing attacks, etc. When a security incident happens and data loss occurs, it is important to evaluate how the lost data affects user privacy. This paper studies the impact of data loss to user privacy in online social networks. We consider two scenarios in the paper, i.e., when a social network user's partial user information is disclosed and when a group of user accounts are compromised. In the first scenario, a social network user is confused by social network privacy settings and has the wrong privacy settings in place. The user's information, such as age, gender, date of birth, and zip code, becomes publicly accessible in the Internet. In the second scenario, a group of user accounts are compromised in a security incident. A malicious user is able to manipulate these compromised user accounts to collect more data from social networks. In either of these two scenarios, it is important to understand how a user's privacy is affected. However, privacy impact assessment is a very challenging issue.

First, lost data may include sensitive personally identifiable information (PII) such as name, user accounts, and birthdate which has a direct impact to privacy. It may also include confidential data, intellectual property and cause financial loss. However, security impact is not limited to the information revealed in the lost data. Deep analytics on the lost data could infer and reveal hidden information. Moreover, huge amount of information is also available via public records, social media, and the Internet. This also leads data leakage. Using the lost data and publicly available information, scams, such as phishing, web scams, and social engineering, could be used to compromise more data. Data loss, indirect data loss, and potential risk due to deep analytics, searchable information, and scams must be assessed and prevented.

Second, most approaches for privacy impact assessment are based on checklists and auditions. There is lack of quantitative analysis approach for privacy impact assessment. Privacy impact assessment requires quantifying and measuring privacy. However, privacy measurement itself is a challenging issue. The definition of privacy is very subjective. People have different opinions about privacy. Privacy can be further described by certain attributes, such as user name, date of birth,

978-1-4673-7648-8/15/$31.00 ©2015 IEEE

and social security number. However, these privacy attributes have different impact to an individual. Furthermore, privacy is also an evolving concept. The Internet has changed many aspects of privacy.

Third, users are all connected in a social network. Lost data may include compromised user accounts. A compromised user account can be further manipulated by a malicious user to collect more information from social network users. Lost data can also be used to start social engineering attacks to compromise more user accounts. Data loss can spread and propagate in the social network following the user relations. Social network service providers often provide user privacy settings to protect user privacy. However, privacy settings are often confusing and not very effective when a user is tagged as a friend [7].

In this paper, we summarize privacy assessment issues and discuss challenges for privacy impact assessment. We define three measurement metrics to assess privacy impact. We further propose a quantitative analysis approach to assess privacy impact for online social networks when data loss occurs. Two particular challenges are considered in the paper, i.e., when a social network user's partial user information is disclosed and when a group of user accounts are compromised. This paper provides a quantitative analysis approach for government agencies, enterprises, and organizations to assess privacy impact for online social networks when a security incident occurs.

The remainder of the paper is organized as follows: Section II discusses the related work. Section III summarizes privacy assessment issues and challenges. Section IV introduces the proposed quantitative analysis approach for privacy impact assessment. Section V summarizes the paper and future works.

II. Related Work

Privacy impact assessment studies how to collet, use, share, and maintain personal identification information. In this paper, we focus on how disclosed information affects a person's privacy in online social networks. A quantitative analysis approach to assess privacy impact requires measuring privacy. There are two main approaches for privacy measurement, i.e., privacy scores and privacy index.

In [8], the authors present an approach in which privacy score is calculated by computing sensitivity and visibility of attributes [8]. Naïve approach for evaluating sensitivity and visibility of attributes is demonstrated in [8]. The authors further extend their works to another approach in [9]. They use Item Response Theory (IRT) to evaluate sensitivity and visibility of attributes when calculating privacy score. The authors use both synthetic and real-world data to show the effectiveness of their approach.

In [10]–[12], the authors propose a SONET model and privacy indexes to measure privacy exposure in online social networks. In the SONET model [10], deep data analytic techniques such as inference and data aggregation are considered. Hidden relations are defined to explore inference information. Virtual attributes are developed to consider data aggregation issues. In [11], [12], privacy indexes and privacy measurement functions are developed to measure privacy in consideration of both attribute sensitivities and visibilities. In [11], the authors consider a single user's privacy exposure issues in a social network. Privacy indexes and privacy measurement functions are developed without considering user relations and user privacy settings. In [12], we extend our work in [11] in consideration of user relations and user privacy settings. Priacy indexes and privacy measuremnt functions are developed to assess privacy exposure between any two users in a social network. The work in [10]–[12] focuses on privacy measurement. This paper focuses on privacy impact assessment. The proposed privacy impact assessment approach in this paper is based on the privacy index and privacy measurement functions proposed in [10]–[12].

A few works attempting to quantify the privacy risk associated with the usage of social networks can be found in the literature. In [13], the authors propose to use the amount of information revealed in online social networks to quantify the privacy risk. However, there is no measurement functions developed in [13]. The authors in [14] develop a tool, Privometer, to measure information leakage. The leakage is measured by a numerical value derived from combined probability of inference. The tool can suggest self-sanitization actions based on the numerical value. In [14], the authors propose to measure the privacy risk based on social networks' privacy policies and practices when handling users [14]. Privacy scores are calculated in a debatable manner. In [15], the authors use risk labelling approach to tag users based on community members' feedback. Active learning method is used to correctly label strangers.

III. Privacy Impact Assessment

A security incident may happen in online social networks and data loss may occur. The impact of lost data to user privacy needs to be evaluated. In this paper, we divide data loss into three categories as shown in Figure 1:

- Direct data loss, direct information revealed in lost data.
- Indirect data loss, data loss due to inferred information, aggregated information, and searchable information.
- Potential risk, data loss due to scams such as phishing, web scams, and social engineering.

A. Data Loss and Its Impact to Privacy

Direct Data Loss: Lost data may include personally identifiable information which will result in direct data loss. PII includes any information which can be used to distinguish of trace an individual's identity such as name, social security number, date and place of birth, mother's maiden name, or biometric records [16]. It also includes other information that is linked or linkable to an individual, such as medical,

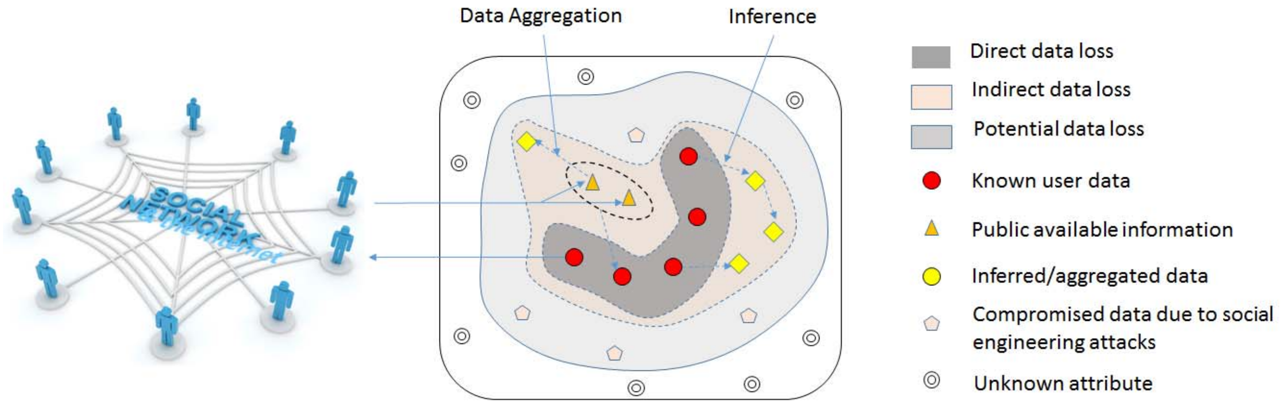

Figure 1. Data Loss and Privacy Impact

educational, financial, and employment information. Data such as social network user profiles and credit card information includes lots of personally identifiable information. PII is often the target of attacks and loss of PII could result in identify theft.

Indirect Data Loss: Huge amount of data is also available through the Internet and public records. These publicly available data is assumed to be known and also accounts for indirect data loss. Deep data analytic technologies such as inference, aggregation, and de-anonymization could be used to obtain more useful data from publicly available information. Social networks include huge amount of user data such as user profiles, relations, and activities. Indirect information could be further inferred from user data. For example, group participations could be used to link to user interests and professions [17]. Certain information could also be combined and used to disclose privacy. For example, the work in [18], [19] found that 87% of Americans can be uniquely identified by five digit zip code, gender, and date of birth. However, none of them alone can significantly affect privacy. In addition, social networking sites sometimes share anonymized user data with third parties. De-anonymization techniques have been found and user identities can be revealed from the anonymized data [20], [21].

Potential Data Loss: Social engineering attacks are one of the main concerns which may cause potential data loss and impact. Using the information retrieved from the lost data, an adversary could craft social engineering attacks pretending to be a genuine friend of target users and convince them to reveal more sensitive information. Studies have found that phishing attempts are more likely to success if an attacker uses the victim's friends' information [22]. Potential data loss and the number of users affected by the lost data might be huge.

B. *Privacy Impact Assessment Challenges*

Most of the research and existing solutions focus on direct data loss and its impact. Indirect data loss, potential data loss and their impact to user privacy must also be considered. However, it is challenging to measure data loss and assess their privacy impact.

First, privacy preserving focuses on protection of PII. However, it is not clear what PII exactly includes. HIPAA Protected Health Information (PHI) lists 18 identifiers which can be used to identify an individual and further specifies the rules to secure PHI in an organization. It is highly possible that there are other identifiers which can be used to link to user identity. For example, the authors found that human mobility traces are highly unique [23]. Using a test dataset where the location of an individual is specified hourly, 95% of the individuals can be uniquely identified using four spatio-temporal points.

Second, the correlations of PII are not clear and it is also uncertain what data should be protected due to deep analytic techniques. Inference attacks exist and can be used to collect more indirect information from existing known data which may be linked to personal identity [10], [17], [24]. Further, Non-sensitive data could be aggregated to reveal more sensitive information and cause identity theft [18], [19].

Third, advanced techniques such as de-anonymization could be used to link anonymous data to personal identity. Many efforts have been conducted on privacy preserving to obscure PII in a data set [25], [26]. However, de-anonymization attacks have been found to re-identify anonymous data [20], [21].

Fourth, numerous public accessible information is available via public records, social media and the Internet. This may lead to more data leakage. Many new technologies have emerged in the last decade, such as social networking sites and search engines. Huge amount of information is available to public. Using these technologies, information which is not available before may become available now.

Fifth, potential data loss may occur due to scams such as phishing, web scams, and social engineering. Social media has become one of the main sources for cybercriminal to collect information and start social engineering attacks [27]. Utilizing

Table 1. Privacy Impact Assessment Metrics

	Direct Loss	Indirect Loss	Potential Loss	Total	Ratio (%)
Num. of user account compromised		0	0		
Num. of user affected (low risk)					
Num. of user affected (high risk)					

social engineering attacks, it allows cybercriminals to breach more user accounts and compromise more data.

C. Privacy Assessment Metrics

This paper targets to estimate how many users are affected when a security incident occurs. We propose to use number of user accounts compromised, number of user affected (low risk), and number of user affected (high risk), and the ration to the total number of users to assess the privacy impact of a security incident (Table 1).

The direct data loss is indicated by the number of user accounts compromised. To estimate the indirect and potential data loss, we use the number of users affected. If a malicious user knows more information about a user after the incident, the user is affected due to the data loss. The risk of the impact can be further divided into two levels:

Low risk: The lost data does not include any PII and the risk of the user to identity theft is low.

High risk: The lost data includes personally identifiable information. The risk of the user to identity theft is high.

We also suggest further identifying the distribution of affected users due to direct data loss, indirect data loss, and potential data loss. The distribution indicates how the data loss and impact spread through the social networks.

IV. A QUANTITATIVE ANALYSIS APPROACH FOR PRIVACY IMPACT ASSESSMENT

We propose a quantitative analysis approach for privacy impact assessment for online social networks. Two particular challenges are addressed in the paper, i.e., when a social network user's partial user information is disclosed, and when a group of social network user accounts are compromised.

A. Privacy Impact Assessment When Partial User Information Is Disclosed

A social network is a network of actors connected together. An actor is a social entity (e.g. people, organization, etc.) in a social network. Let A_i and A_j be two actors in the social network. A_j's privacy exposure to A_i is decided by the following factors:

- **$A_j's$ user profile**: An actor has certain characteristics that describe its features known as attributes. An attribute may or may not be available in A_j's user profile. Let $L = \{a_1, a_2, \dots, a_n\}$ represent an actor's attributes ($a_i = 0\ or\ 1$). We use $A_i(a_{i1}, a_{i2}, \dots, a_{in})$ and $A_j(a_{j1}, a_{j2}, \dots, a_{jn})$ to represent A_i and A_j's user profiles.

- **Attribute sensitivity**: Each attribute has a different impact on privacy. This impact is referred as Attribute Privacy Impact Factor (APIF). We use $S = \{s_1, s_2, \dots, s_n\}, 0 \le s_i \le 1$ to represent the corresponding privacy impact factors of these attributes.

- **A_i and A_j's user relationships**: the user relations are decided by the degree of separation between A_i and A_j. Friend relations in a social network can be usually divided into three groups, e.g., friends (1 degree actors), friends of friends (2 degree actors), public (3 and above degree actors). We use degrees of separation function h

$$d_{ij} = h(A_i, A_j)$$

to represent the distance between actor A_i and A_j.

- **$A_j's$ user privacy settings**: privacy settings decide if an attribute is visible to a friend. Privacy settings can be defined using the degree of separation. $A_j's$ privacy settings to A_i can be represented by

$$P_j(i) = \{p_{j1}(i), p_{j2}(i), \cdots, p_{jn}(i)\}$$

where $p_{jt}(i) = 0\ or\ 1\ (1 \le t \le n)$ and

$$p_{jt}(i) = g_j(a_{jt}, d_{ij})$$

g_j is A_j's attribute visibility function.

The privacy impact due to partial user information loss can be evaluated using privacy index and privacy measurement functions proposed in [3] [4]. We briefly introduce the privacy index and the privacy measurement functions in this section.

Privacy Index *PIDX(i,j)* was proposed to describe actor $\mathrm{A_j}$'s privacy exposure to $\mathrm{A_i}$ based on $\mathrm{A_j}'$s visible attributes to $\mathrm{A_i}$. High *PIDX* value indicates high exposure of privacy. Privacy Index *PIDX* is between 0 and 100.

$$PIDX(i,j) = \frac{w(i,j)}{w(j)} \times 100$$

where $w(i,j)$ represents the privacy weight of $A_j's$ visible attributes to A_i and $w(j)$ to represent the maximum privacy weight of $A_j's$ attributes. Let f be a privacy measurement function which returns a numeric value on L_j, S, $P_j(i)$. We have

$$w(i,j) = f(L_j, S, P_j(i))$$

According to our definition, we have $p_{jt}(j) = 1$ (A user's data is always available to him/herself.) and $w(j,j)$ returns the maximum privacy weight of actor A_j. Thus,

$$w(j) = w(j,j) = f(L_j, S, P_j(j))$$

We also have $PIDX(i) = PIDX(i,i)$ which could be used to measure user $i's$ privacy exposure in a social network.

Three privacy measurement functions were proposed to measure A_j's privacy exposure to A_i in [12], i.e., weighted privacy measurement function, maximum privacy measurement function, and composite privacy measurement function. Three privacy indexes are defined accordingly.

Weighted privacy measurement function is defined as

$$f_w\left(L_j, S, P_j(i)\right) = s_1 p_{j1}(i) + s_2 p_{j2}(i) + \cdots + s_n p_{jn}(i) = \sum_{t=1}^{n} s_t p_{jt}(i)$$

w-PIDX(i,j) is an index which measures actor $A_j's$ privacy exposure to A_i.

$$w - PIDX(i,j) = \frac{w(i,j)}{w(j)} \times 100 = \frac{f\left(L_j, S, P_j(i)\right)}{f\left(L_j, S, P_j(j)\right)} \times 100 = \frac{\sum_{t=1}^{n} s_t g_j(a_{jt}, d_{ij})}{\sum_{t=1}^{n} s_j} \times 100$$

Maximum privacy measurement function is defined as

$$f_m\left(L_j, S, P_j(i)\right) = \max(s_1 p_{j1}(i), s_2 p_{j2}(i), \cdots, s_n p_{jn}(i))$$

where *max* is a function returning the maximum value in the list.

m-PIDX(i,j) is an index which measures actor $A_j's$ maximum privacy exposure to A_i

$$m - PIDX(i,j) = f\left(L_j, S, P_j(i)\right) \times 100 = max(s_1 g_j\left(a_{j1}, d_{ij}\right), s_2 g_j(a_{j2}, d_{ij}) \ldots, s_n g_j(a_{jn}, d_{ij})) \times 100$$

Composite privacy measurement function is defined as

$$f_c\left(L_j, S, P_j(i)\right) = f_m\left(L_j, S, P_j(i)\right) + \left(1 - f_m\left(L_j, S, P_j(i)\right)\right) \times \frac{f_w\left(L_j, S, P_j(i)\right)}{f_w\left(L_j, S, P_j(j)\right)}$$

c-PIDX(i,j) is an index which measures actor $A_j's$ privacy exposure to A_i based on $A_j's$ composite privacy measurement function. *c-PIDX(i,j)* is defined as

$$c - PIDX(i,j) = f_c\left(L_j, S_j, P_j(i)\right) \times 100$$

c-PIDX(i,j) can be represented using *w-PIDX* and *m-PIDX* as below:

$$c - PIDX(i,j) = m - PIDX(i,j) + (100 - m - PIDX(i,j)) * \frac{w - PIDX(i,j)}{100}$$

As found in [4], *w-PIDX(i,j)* is good at reflecting attribute incremental changes. However, *w-PIDX(i,j)* does not reflect the actual privacy exposure. *m-PIDX(i,j)* can be used for privacy ranking. However, *m-PIDX(i,j)* does not reflect the attribute incremental changes. *c-PIDX(i, j)* is a good indication of actor $A_j's$ privacy exposure to A_i. Note that *c-PIDX(i, j)* might not equal *c-PIDX(j,i)* because A_i and A_j may have different privacy settings. We use *c-PIDX(i, j)* for privacy measurement in social networks and *c-PIDX(i,j)* is also the default privacy measurement function for *PIDX(i,j)*.

B. Privacy Impact Assessment When a Group of User Accounts Are Compromised

In case a group of user accounts are compromised, it is important to assess the impact of the incidents to others. Using the privacy index and privacy measurement functions developed, we are able to assess privacy impact when a group of user accounts are compromised.

Let $G = (V, E)$ represent a social network graph and V_S be a set of victims whose accounts are compromised. A malicious user *m* could masquerade as a victim *v* and convince *v's* friends to steal their personal data. Since users are connected together through friend list, the incident may spread through the friend list and affect more people. We use a threshold T to differentiate users affected with low and high risk.

Users not affected, users are not affected if their privacy exposure does not change before and after the incident.

Users affected with low risk, a malicious user knows more about a target after the incident. Let PEI be the privacy exposure after the incident. We have $PEI \leq T$. If PEI is less than T, the user is at low risk and should not be affected by the incident.

Users affected with high risk, a malicious user knows more about a target after the incident and the privacy index is higher than the threshold T. The user is at high risk and is possibly affected by the incident.

Assume the user profile, data sensitivity, user relations, privacy settings are available. The algorithm shown in Figure 2 returns the number of users who are *d* steps away from the victims. Q is a FIFO queue in the algorithm. $h\left(A_i, A_j\right)$ returns the degree separation of actors A_i and A_j. The algorithm returns an expanded user set V_E which includes all the users having $PEI \geq T$ and within the *d* steps away from a victim in V_S.

C. Privacy Impact Assessment Analysis

```
for each actor A in $V_S$
    if $PIDX(A) \geq T$  add A to $V_E$;

    add $A$'s friends to Q;
    for each actor F in Q
        if $h(A, F)+1 < d$ then
            add F's friends to Q;
            if $PIDX(A, F) \geq T$  then
                add F to $V_E$;
            end If
        end if
    end for
end for
```

Figure 2. Number of Users Who Are *d* Steps Away from the Victims

A social network user's privacy is decided by what data is available in the social network (user profile), the sensitivity of the user data, the user relations, and the user privacy settings. Data sensitivity is decided by how an attribute links to user identity. Data availability is decided by user profile, user relations, and user privacy settings. The privacy index and the privacy measurement functions consider both attribute sensitivity and availability.

The proposed approach can be used for privacy impact assessment when partial user information is disclosed or when a group of user accounts are compromised. Indirect data loss and impact due to inference and data aggregation can be assessed too. However, indirect data loss due to de-anonymization and potential data loss due to social engineering attacks are not considered in the paper.

V.CONCLUSIONS AND FUTURE WORKS

In this paper, we propose a quantitative analysis approach for privacy impact assessment. Our focus is to assess how user's privacy is affected when a social network user's partial user information is disclosed or when a group of social network user accounts are compromised. The proposed approach requires using a privacy measurement function to measure user *j's* privacy exposure to user *i*. We use the privacy index and the privacy measurement functions proposed in [11], [12] to measure the privacy exposure.

The proposed approach could be used to assess the privacy impact due to direct data loss and indirect data loss (inference information, aggregation information). Impact to privacy due to potential data loss might be supported too. However, more studies need to be conducted to study how malicious users use collected data set for social engineering attacks. Our future work also includes theoretic analysis and study on the relations between the number of user affected and the number of user accounts compromised.

REFERENCES

[1] Facebook, "Key Facts," 2014. [Online]. Available: http://newsroom.fb.com/Key-Facts. [Accessed: 30-Jun-2014].

[2] M. W. DiStaso, "A Benchmark Analysis of the Strategic Use of Social Media for Fortune's Most Admired U.S. Companies on Facebook, Twitter and Youtube," *Public Relat. J.*, vol. 7, no. 1, pp. 1–33, 2013.

[3] E. Lee, "Associated Press Twitter Account Hacked in Market-Moving Attack," *Bloomberg Technology*, 24-Apr-2013.

[4] McAfee, "Securely Enabling Social Media," 2013.

[5] Verizon, "2013 Data Breach Investigations Report," 2013.

[6] Checkpoint, "The Risk of Social Engineering on Information Security: A Survey of IT Professionals," 2011.

[7] Y. Liu, K. P. Gummadi, and A. Mislove, "Analyzing Facebook Privacy Settings : User Expectations vs . Reality," in *IMC' 11*, 2011.

[8] E. M. Maximilien, T. Grandison, T. Sun, D. Richardson, S. Guo, and K. Liu, "Privacy-as-a-service: Models, algorithms, and results on the facebook platform," in *Proceedings of Web*, 2009, vol. 2.

[9] K. Liu, "A Framework for Computing the Privacy Scores of Users in Online Social Networks," *Knowl. Discov. Data*, vol. 5, no. 1, pp. 1–30, 2010.

[10] R. K. Nepali and Y. Wang, "SONET: A SOcial NETwork Model for Privacy Monitoring and Ranking," in *The 2nd International Workshop on Network Forensics, Security and Privacy*, 2013.

[11] Y. Wang and R. N. Kumar, "Privacy Measurement for Social Network Actor Model," in *The 5th ASE/IEEE International Conference on Information Privacy, Security, Risk and Trust*, 2013.

[12] Y. Wang, R. K. Nepali, and J. Nikolai, "Social Network Privacy Measurement and Simulation," in *2014 International Conference on Computing, Networking and Communication (ICNC), CNC Workshop*, 2014.

[13] J. Becker and H. Chen, "Measuring Privacy Risk in Online Social Networks," in *Web 2.0 security and privacy Workshop*, 2009.

[14] N. Talukder, M. Ouzzani, A. K. Elmagarmid, H. Elmeleegy, and M. Yakout, "Privometer: Privacy protection in social networks," in *2010 IEEE 26th International Conference on Data Engineering Workshops (ICDEW 2010)*, 2010, pp. 266–269.

[15] C. Akcora, B. Carminati, and E. Ferrari, "Privacy in Social Networks: How Risky is Your Social Graph?," in *2012 IEEE 28th International Conference on Data Engineering*, 2012, pp. 9–19.

[16] NIST SP 800-122, *Guide to Protecting the Confidentiality of Personally Identifiable Information*. 2010, p. 59.

[17] E. Zheleva and L. Getoor, "To join or not to join : The illusion of privacy in social networks with mixed public and private user profiles," in *WWW 2009*, 2009, pp. 531–540.

[18] L. Sweeney, "Uniqueness of Simple Demographics in the U. S. Population," in *Data privacy Lab white paper series LIDAP-WP4*, 2000.

[19] P. Golle, "Revisiting the uniqueness of simple demographics in the US population," in *Proceedings of the 5th ACM workshop on Privacy in electronic society*, 2006, pp. 77–80.

[20] A. Narayanan and V. Shmatikov, "De-anonymizing Social Networks," in *2009 30th IEEE Symposium on Security and Privacy*, 2009, pp. 173–187.

[21] G. Wondracek, T. Holz, E. Kirda, and C. Kruegel, "A Practical Attack to De-anonymize Social Network Users," *2010 IEEE Symp. Secur. Priv.*, pp. 223–238, 2010.

[22] T. Jagatic, N. Johnson, M. Jakobsson, and F. Menczer, "Social Phishing," *Commun. ACM*, vol. 50, no. 10, pp. 94–100, 2007.

[23] Y.-A. de Montjoye, C. a Hidalgo, M. Verleysen, and V. D. Blondel, "Unique in the Crowd: The privacy bounds of human mobility.," *Sci. Rep.*, vol. 3, p. 1376, 2013.

[24] J. Tang, T. Lou, and J. Kleinberg, "Inferring social ties across heterogeneous networks," in *WSDM'12*, 2012, pp. 743–752.

[25] L. Sweeney, "K-anonymity: a Model For Protecting Privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05. pp. 557–570, 2002.

[26] A. Machanavajjhala, J. Gehrke, D. Kifer, and M. Venkitasubramaniam, "L-Diversity: Privacy beyond k-anonymity," in *Proceedings - International Conference on Data Engineering*, 2006, vol. 2006, p. 24.

[27] D. Dieterle, "Hackers Target Social Media for Social Engineering Attacks," *Infosec Island*, Mar-2012.