3-2015

# Privacy leakage analysis in online social networks

Yan LI
*Singapore Management University*, liyan@smu.edu.sg

Yingjiu LI
*Singapore Management University*, yjli@smu.edu.sg

Qiang YAN
*Singapore Management University*, qiang.yan.2008@smu.edu.sg

DENG, Robert H.
*Singapore Management University*, robertdeng@smu.edu.sg

Published in Computers & Security, 2015 March, Volume 49, pp. 239-254.
http://doi.org/10.1016/j.cose.2014.10.012

# Privacy leakage analysis in online social networks

*Yan Li*[*], *Yingjiu Li*, *Qiang Yan*, *Robert H. Deng*

*School of Information Systems, Singapore Management University, Singapore*

*Keywords:*
Online social network
Privacy control
Information flow
Private information leakage
Inherent exploit

A B S T R A C T

Online Social Networks (OSNs) have become one of the major platforms for social interactions, such as building up relationship, sharing personal experiences, and providing other services. The wide adoption of OSNs raises privacy concerns due to personal data shared online. Privacy control mechanisms have been deployed in popular OSNs for users to determine who can view their personal information. However, user's sensitive information could still be leaked even when privacy rules are properly configured. We investigate the effectiveness of privacy control mechanisms against privacy leakage from the perspective of information flow. Our analysis reveals that the existing privacy control mechanisms do not protect the flow of personal information effectively. By examining representative OSNs including Facebook, Google+, and Twitter, we discover a series of privacy exploits. We find that most of these exploits are inherent due to the conflicts between privacy control and OSN functionalities. The conflicts reveal that the effectiveness of privacy control may not be guaranteed as most OSN users expect. We provide remedies for OSN users to mitigate the risk of involuntary information leakage in OSNs. Finally, we discuss the costs and implications of resolving the privacy exploits.

## 1. Introduction

Online Social Networks (OSNs) have become an essential element in modern life for human beings to stay connected to each other. About 82% online population use at least one OSN such as Facebook, Google+, Twitter, and Linkedln, which facilitates building relationship, sharing personal experiences, and providing other services (Aquino, 2012). Via OSNs, massive amount of personal data is published online and accessed by users from all over the world. Prior research (Zheleva and Getoor, 2009; Johnson et al., 2012; Chaabane et al., 2012; Balduzzi et al., 2010) shows that it is possible to infer undisclosed personal data from *publicly* shared information. Nonetheless, the availability and quality of the public data causing privacy leakage are decreasing due to the following reasons: 1) privacy control mechanisms have become the standard feature of OSNs and keep evolving. 2) the percentage of users who choose *not* to publicly share information is also increasing (Chaabane et al., 2012). In this tendency, it seems that privacy leakage could be *prevented* as increasingly comprehensive privacy control is in place. However, this may not be achievable according to our findings.

Instead of focusing on new attacks, we investigate the problem of *privacy leakage under privacy control* (PLPC). PLPC refers to private information leakage even if privacy rules are

* *Corresponding author.*
E-mail addresses: yan.li.2009@smu.edu.sg (Y. Li), yjli@smu.edu.sg (Y. Li), qiang.yan.2008@smu.edu.sg (Q. Yan), robertdeng@smu.edu.sg (R.H. Deng).

properly configured and enforced. For example, Facebook allows its users to control over who can view their friend lists on Facebook. Alice, who has Bob in her friend list on Facebook, may not allow Bob to view her complete friend list. As an essential functionality, Facebook recommends to Bob a list of users, called "*people you may know*", to help Bob make more friends. This list is usually compiled by enumerating the friends of Bob's friends on Facebook, which includes Alice's friends. Even though Alice doesn't allow Bob to view her friend list, Alice's friend list could be leaked as recommendation to Bob by Facebook.

We investigate the underlying reasons that make privacy control vulnerable from the perspective of information flow. We start with categorizing the personal information of an OSN user into three *attribute sets* according to *who the user is*, *whom the user knows*, and *what the user does*, respectively. We model the information flow between these attribute sets and examine the functionalities which control the flow. We inspect representative real-world OSNs including Facebook, Google+, and Twitter, where privacy exploits and their corresponding attacks are identified.

Our analysis reveals that most of the privacy exploits are inherent due to the underlying conflicts between privacy control and essential OSN functionalities. The recommendation feature for social relationship is a typical example, where it helps expanding a user's social network but it may also conflict with other users' privacy concerns for hiding their social relationships. Therefore, the effectiveness of privacy control may not be guaranteed even if it is technically achievable. We investigate necessary conditions for protecting against privacy leakage due to the discovered exploits and attacks. Based on the necessary conditions, we provide suggestions for users to minimize the risk of involuntary information leakage when sharing private personal information in OSNs.

We analyze the potentially vulnerable users due to our identified attacks through user study, in which we investigate participants' usage, knowledge, and privacy attitudes towards Facebook, Google+, and Twitter. Based on the collected data, we investigate the vulnerability of these participants who could leak the private information through the attacks. We further discuss the costs and implications of resolving these privacy exploits.

We summarize the contributions of this paper as follows:

- We investigate the interaction between privacy control and information flow in OSNs. We show that the conflict between privacy control and essential OSN functionalities restricts the effectiveness of privacy control in OSNs.
- We identify privacy exploits for current privacy control mechanisms in typical OSNs, including Facebook, Google+, and Twitter. Based on these privacy exploits, we introduce a series of attacks for adversaries with different capabilities to obtain private personal information.
- We investigate necessary conditions for protecting against privacy leakage due to the discovered exploits and attacks. We provide suggestions for users to minimize the risk of privacy leakage in OSNs. We also analyze the costs and implications of resolving discovered exploits. While it is possible to fix the exploits due to implementation defects, it is not easy to eliminate the inherent exploits due to the conflicts between privacy control and the functionalities. These conflicts reveal that the effectiveness of privacy control may not be guaranteed as most OSN users expect.

The rest of this paper is organized as follows: Section 2 provides background information about OSNs. Section 3 presents our threat model and assumptions. Section 4 models information flows between attribute sets in OSNs. Section 5 presents discovered exploits, attacks, and mitigations for the exploits. Section 6 analyzes the potentially vulnerable users due to the attacks. Section 7 discusses the implications of our findings. Finally, Section 8 describes related work and Section 9 summarizes our conclusions.

## 2. Background

In a typical OSN, Alice owns a space which consists of a *profile page* and a *feed page* for publishing Alice's personal information and receiving other users' personal information, respectively. Alice's profile page displays Alice's personal information, which can be viewed by others. Alice's feed page displays other users' personal information which Alice would like to keep up with. The personal information in a user's profile page can be categorized into three *attribute sets*: a) personal particular set (*PP* set), b) social relationship set (*SR* set), and c) social activity set (*SA* set), according to who the user is, whom the user interact with, and what the user does, respectively. We show corresponding personal information and attribute sets on Facebook, Google+, and Twitter in Table 1.

Alice's *PP* set describes persistent facts about Alice in an OSN, such as gender, date of birth, and race, which usually do not change frequently. Alice's *SR* set records her social relationships in an OSN, which consist of an *incoming list* and an *outgoing list*. The incoming list consists of the users who include Alice as their friends while the outgoing list consists of the users whom Alice includes as her friends. In particular, on Google+, the incoming list and the outgoing list correspond to "have you in circles" and "your circles", respectively. On Twitter, the incoming list and the outgoing list correspond to "following" and "follower", respectively. The social relationships in certain OSNs are mutual. For example, on Facebook, if Alice is a friend of Bob, Bob is also a friend of Alice. In such a case, a user's incoming list and outgoing list are the same, which are called friend list. Lastly, Alice's *SA* set describes Alice's social activities in her daily life. The *SA* set includes status messages, photos, links, videos, etc.

To enable users protect their personal information in the three attribute sets, most OSNs provide privacy control, by which users may set up certain *privacy rules* to control the disclosure of their personal information. Given a piece of personal information, the privacy rules specify who can/cannot view the information. A privacy rule usually contains two types of lists, *white list*, and *black list*. A white list specifies who can view the information while a black list specifies who cannot view the information. A white/black list could be local or global. If a white/black list is local, this list takes effect on specific information only (e.g. an activity, age information, or

**Table 1 – Types of personal information on Facebook, Google+, and Twitter.**

| Acronym | Attribute set | Facebook | Google+ | Twitter |
|---|---|---|---|---|
| PP | Personal particulars | Current city, hometown, sex, birthday, relationship status, employer, college/university, high school, religion, political views, music, books, movies, emails, address, city, zip | Taglines, introduction, bragging rights, occupation, employment, education, places lived, home phone, relationship, gender | Name, location, bio, website |
| SR | Social relationship (incoming list, outgoing list) | Friends, friends | Have you in circles, your circles | Following, follower |
| SA | Social activities | Status message, photo, link, video, comments, like | Post, photo, comments, link, video, plus 1's | Tweets |

gender information). If a white/black list is global, this list takes effect on all information in a user's profile page. For example, if Alice wants to share a status with all her friends except Bob, Alice may use a local white list which includes all Alice's friends, as well as a local black list which includes Bob only. If Alice doesn't want to share any information with Bob, she may use a global black list which includes Bob.

To help users share their personal information and interact with each other, most OSNs provide four basic functionalities including *PUB*, *REC*, *TAG*, and *PUSH*. The first three functionalities, *PUB*, *REC*, and *TAG*, mainly affect the personal information displayed in a user's profile page, while the last functionality *PUSH* makes some other users' personal information appear in the user's feed page. These basic functionalities are described as follows. We exclude any other functionalities which are not relevant to our findings.

Alice can use *PUB* functionality to share her personal information with other users. As shown in Fig. 1a, *PUB* displays Alice's personal information in her profile page. Other users may view Alice's personal information in Alice's profile page.

To help Alice make more friends in an OSN, *REC* is an essential functionality by which the OSN recommends to Alice a list of users that Alice may include in her *SR* set. The list of recommended users is composed based on the social relationships of the users in Alice's *SR* set. Considering an example shown in Fig. 1b, Alice's *SR* set consists of Bob while Bob's *SR* set consists of Alice, Carl, Derek, and Eliza. After Alice logs into her space, *REC* automatically recommends Carl, Derek, and Eliza to Alice who may update her *SR* set. If Alice intends to include Carl in her *SR* set, Alice may need Carl's approval depending on OSN implementations. Upon approval if needed, Alice can include Carl in her *SR* set. At the same time, Alice is automatically included in Carl's *SR* set. In particular, on Facebook, if Alice intends to include Carl in her *SR* set, Alice needs to get Carl's approval. Upon approval, Alice includes Carl in her friend list. Meanwhile, Facebook automatically includes Alice in Carl's friend list. On Google+, Alice can include Carl in her outgoing list without Carl's approval. Then Google+ automatically includes Alice in Carl's incoming list. On Twitter, if Alice intends to include Carl in her *SR* set, Alice may need Carl's approval depending on Carl's option whether his approval is required. Upon approval if required, Alice includes Carl in her incoming list. Then Twitter includes Alice in Carl's outgoing list automatically.

To motivate users' interactions, *TAG* functionality allows a user to mention another user's name in his/her social activities when the user publishes social activities in his/her profile page. In Fig. 1c, when Alice publishes a social activity in her profile page, she can mention Bob in the social activity via *TAG*, which provides a link to Bob's profile page (shown as an HTML hyperlink).

For the convenience of keeping up with the personal information published by other users, OSNs provides *feed page* for users. Considering an example in which Alice intends to keep up with Bob, Alice can subscribe to Bob, and Alice is called Bob's *subscriber*. As Bob's subscriber, Alice is included in Bob's *SR* set. In particular, on Facebook, a user's subscribers are usually his/her "friends". On Google+, a user's subscribers are usually the users in his/her outgoing list, i.e. "your circles". On Twitter, a user's subscribers are usually the users in his/her

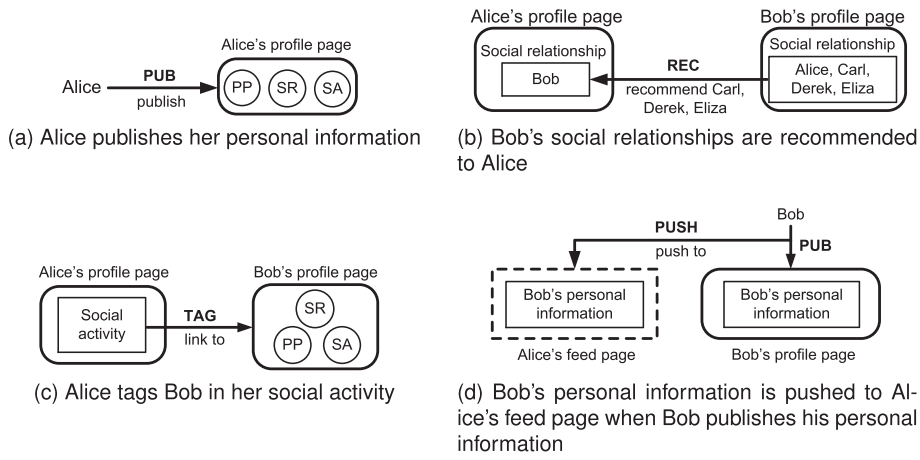


**Fig. 1 – Basic functionalities in OSNs.**

incoming list, i.e. "follower". Fig. 1d shows that when Bob updates his personal information via *PUB* and allows Alice to view the updated personal information, a copy of the updated personal information is automatically pushed to Alice's feed page via *PUSH*. Then, Alice can view Bob's updated personal information both in her feed page and in Bob's profile page.

## 3. Threat model

The problem of PLPC investigates privacy leakage in a system where privacy control is enforced. Given a privacy control mechanism, *PLPC* examines whether a user's private personal information is leaked *even if* the user properly configures privacy rules to protect the corresponding information.

The problem of PLPC in OSNs involves two parties, *distributor* and *receiver*. A user who publishes and shares his/her personal information is a *distributor* while the user whom the personal information is shared with is a *receiver*. An *adversary* is a receiver who intends to learn a distributor's information that is not shared with him. Correspondingly, the target distributor is referred to as *victim*.

Prior research (Zheleva and Getoor, 2009; Chaabane et al., 2012; Balduzzi et al., 2010) mainly focuses the inference of undisclosed user information from their *publicly* shared information. Since the effectiveness of these inference techniques will be hampered by increasing user awareness of privacy concern (Chaabane et al., 2012), we further include *insiders* in our analysis. The adversaries have the incentive to register as OSN users so that they may directly access a victim's private personal information or infer the victim's private personal information from other users connected with the victim in OSNs.

The capabilities of an adversary can be characterized according to two factors. The first factor is the distance between adversary and victim. According to privacy rules available in existing OSNs, a distributor usually chooses specific receivers to share her information based on the distance between the distributor and the receivers. Therefore, we classify an adversary's capability based on his distance to a victim. Considering the social network as a directed graph, the distance between two users can be measured by the number of hops in the shortest connected path between the two users. An $n$-hop adversary can be defined such that the length of the shortest connected path from victim to adversary is $n$ hops. We consider the following three types of adversaries in our discussion, 1-hop adversary, 2-hop adversary, and $k$-hop adversary, where $k > 2$. On Facebook, they correspond to Friend-only, Friend-of-Friend, and Public, respectively. On Google+, they correspond to Your-circles, Extended-circles, and Public, respectively. For ease of readability, we use *friend*, *friend of friend*, and *stranger* to represent 1-hop adversary, 2-hop adversary, and $k$-hop adversary (where $k > 2$) adversaries, respectively: 1) If an adversary is a friend of a victim, he is stored in the outgoing list in the victim *SR* set. The adversary can view the victim's information that is shared with her friends, friends of friends, or all receivers in an OSN. However, the adversary cannot view the information that is not shared with any receivers (e.g. the "only me" option on Facebook). 2) If an adversary is a friend of friend, he can view the victim's information shared with her friend-of-friends or all receivers. However, the adversary cannot view any information that is shared with friends only, or any information that is not shared with any receivers. 3) If an adversary is a stranger, he can access the victim's information that is shared with all receivers. However, the adversary cannot view any information which is shared with friends of friends and friends.

Besides the above restrictions, an adversary cannot view a victim's personal information if the adversary is included in the victim's black lists (e.g. "except" or "block" option on Facebook, and "block" option on Google+).

An adversary may have prior knowledge about a victim. We will specify the exact requirement of such prior knowledge for different attacks in Section 5.

Since a user may use multiple OSNs, it is possible for an adversary to infer the user's private data by collecting and analyzing the information shared in different OSNs. We exclude social engineering attacks where a victim is deceived to disclose her private information voluntarily. We also exclude privacy leakage caused by improper privacy settings. These two cases cannot be addressed completely by any technical measures alone.

## 4. Information flows between attribute sets in profile pages

In this section, we examine explicit and implicit information flows in OSNs. These information flows could leak users' private information to an adversary even after the users have properly configured the privacy rules to protect their information.

As analyzed in Section 2, the personal information shared in a user's profile page can be categorized into three attribute sets including *PP* set, *SR* set, and *SA* set, which are illustrated as circles in Fig. 2. The attribute sets of multiple users are connected within an OSN, where personal information may explicitly flow from a profile page to another profile page via *inter-profile functionalities*, including *REC* (recommending) and *TAG* (tagging), as represented by solid arrows and rectangles in Fig. 2. It is also possible to access a user's personal information in *PP* set and *SR* set via implicit information flows marked by dashed arrows. The details about these information flows are described below.

The first explicit flow is caused by *REC*, as shown in arrow (1) in Fig. 2. *REC* recommends to an OSN user Bob a list of users according to the social relationships of the users included in Bob's *SR* set. Therefore, the undisclosed users included in Alice's *SR* may be recommended to Bob via *REC*, if Bob is connected with Alice.

The second explicit flow caused by *TAG* is shown in arrow (2) in Fig. 2. A typical OSN user may mention the names of other users in a social activity in *SA* set in his/her profile page via *TAG*, which creates explicit links connecting *SA* sets within different profile pages.

The third flow is an implicit flow caused by the design of information storage for *SR* sets, which is shown in arrow (3) in Fig. 2. A user's *SR* set stores his/her social relationships as connections. From the perspective of information flow, a *connection* is a directional relationship between two users, including a *distributor* and his/her *1-hop receiver*, i.e. *friend*. The direction of a connection represents the direction of information flow. Correspondingly, Alice's *SR* set consists of an *incoming list* and an *outgoing list* as defined in Section 2. For each user $u_i$ in Alice's incoming list, there is a connection from $u_i$ to Alice. For each user $u_o$ in Alice's outgoing list, there a connection from Alice to $u_o$. Alice can receive information distributed from the users in her incoming list, and distribute her information to the users in her outgoing list. Given a connection from Alice to Bob, Bob is included in the outgoing list in Alice's *SR* set. Meanwhile Alice is included in the incoming list in Bob's *SR* set. The social relationships in certain OSNs such as Facebook are mutual. Such mutual relationship can be considered as a pair of connections linking two users with opposite directions, similar to replacing a bidirectional edge with two equivalent unidirectional edges.

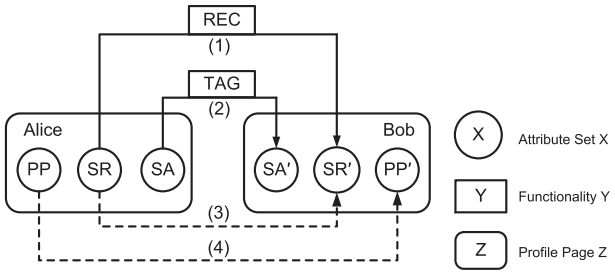


**Fig. 2 – Information flows between attribute sets.**

The fourth flow is an implicit flow related to *PP* set, which is shown as the arrow (4) in Fig. 2. Due to the homophily effect (McPherson et al., 2001; Centola et al., 2007), a user is more willing to connect with the users with similar personal particulars compared to other users with different personal particulars. This tendency can be used to link *PP* sets of multiple users. For example, colleagues working in the same department are often friends with each other on Facebook.

In addition to the above information flows, an OSN user may simultaneously use multiple OSNs, and thus create other information flows connecting the attribute sets of the same user across different OSNs.

It is difficult to prevent privacy leakage from all these information flows. A user may be able to prevent privacy leakage caused by explicit information flows by carefully using corresponding functionalities, as these flows are materialized only when inter-profile functionalities are used. However, it is difficult to avoid privacy leakage due to implicit information flows, as they are caused by inherent correlations among the information shared in OSNs. In fact, all these four information flows illustrated in Fig. 2 correspond to inherent exploits, which will be analyzed in Sections 5 and 7. The existence of these information flows introduces a large attack surface for an adversary to access undisclosed personal information if any of these flows is not properly protected. The existing privacy control mechanisms (Carminati et al., 2009; Fong et al., 2009) regarding data access within a profile page are not sufficient to prevent against privacy leakage. However, the full coverage of privacy control may not be feasible as it conflicts with social/business values of OSNs as analyzed in Section 7.

In this paper, we focus on the information flows from the attribute sets in a profile page to the attribute sets in another profile page, which may lead to privacy leakage even if users properly configure their privacy rules. There may exist other exploitable information flows leading to privacy leakage, which are left as our future work.

## 5. Exploits, attacks, and mitigations

In this section, we analyze the exploits and attacks which may lead to privacy leakage in existing OSNs even if privacy controls are enforced. We organize the exploits and attacks according to their targets, which could be a victim's *PP* set, *SR* set, and *SA* set. We also investigate necessary conditions regarding prevention of privacy leakage due to the identified exploits and attacks. The proofs of the necessary conditions are available in Appendices. Based on these necessary conditions, we provide suggestions on mitigating the corresponding exploits and attacks. All of our findings have been verified in real-world settings on Facebook, Google+, and Twitter.[1]

[1] All of our experiments were conducted from September, 2011 to September, 2012.

### 5.1. *PP set*

A user's *PP* set describes persistent facts about who the user is. The undisclosed information in *PP* set protected by existing privacy control mechanisms can be inferred by the following inherent exploits, namely *inferable personal particular* and *cross-site incompatibility*.

#### 5.1.1. *Inferable personal particular*

Human beings have the tendency to interact with others who share the same or similar personal particulars (such as race, organization, and education). This assumption is called homophily (McPherson et al., 2001; Centola et al., 2007). Due to homophily, users are connected with those who have similar personal particulars at higher rate than with those who have dissimilar personal particulars. This causes an inherent exploit named *inferable personal particulars*, which corresponds to the information flow shown as dashed arrow (4) in Fig. 2.

**Exploit 1**. *If most of a victim's friends have common or similar personal particulars (such as employer information), it could be inferred that the victim may have the same or similar personal particulars.*

An adversary may use Exploit 1 to obtain undisclosed personal particulars in a victim's *PP* set. The following is a typical attack on Facebook.

**Attack 1**. Considering a scenario on Facebook shown in Fig. 3, where Bob, Carl, Derek, and some other users are Alice's friends, and Bob is a friend of Carl, Derek, and most of Alice's friends (Note that in Fig. 3, a solid arrow connects from a distributor to a friend of the distributor). Alice publishes her employer information "*XXX Agency*" in her PP set and allows Carl and Derek only to view her employer information. However, most of Alice's friends may publish their employer information and allow their friends to view this information due to different perceptions in privacy protection. In this setting, Bob can collect the employer information of Alice's friends and infer that Alice's employer is "*XXX Agency*" with high probability.

The above attack works on Facebook, Google+, and Twitter. The attack can be performed by any adversary who has two types of knowledge. The first type of knowledge includes a large portion of users stored in the victim's *SR* set. The second type of knowledge includes the personal particulars of these users. The attack may lead to the leakage of the personal particulars including employer, university, current city, religion, etc. Other personal particulars, including gender, age, and relationship status, may not be reliably inferred via the above attack. But they can be leaked based on additional context information such as users' and their friends' names, social activities, and interests (Tang et al., 2011). To prevent against privacy leakage due to Exploit 1, the following necessary condition should be satisfied.

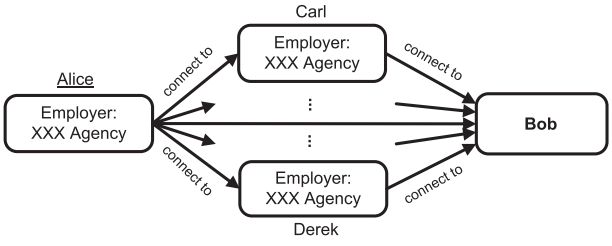


**Fig. 3 – Alice and most of her friends have common personal particulars (e.g. employer information).**

**Necessary Condition 1**. *Given a subset* $U = \{u_1,u_2,\ldots,u_n\}$ *of a victim* $v$*'s SR set in an OSN and personal particular value* $pp_{u_i}$ $(pp_{u_i} \neq null)$ *of each receiver* $u_i \in U$ *which are obtained by an adversary, there exists at least one personal particular value pp such that* $|U_{pp}| \geq |U_v|$ *and* $pp \neq pp_v$ *where* $pp_v$ *is the victim's personal particular value and* $U_{pp} = \{u_i \,|\, (u_i \in U) \wedge (pp_{u_i} = pp)\}$ *and* $U_v = \{u_j \,|\, (u_j \in U) \wedge (pp_{u_j} = pp_v)\}$.

To satisfy Necessary Condition 1, the following mitigations are suggested.

**Mitigation 1**. *If a victim publishes information in her PP set and allows a set of receivers to view the information, the privacy rules chosen by the victim should be propagated to all users in the victim's SR set who have similar or common information in their PP sets.*

**Mitigation 2**. *A victim should intentionally set up a certain number of connections with other users who have different personal particulars.*

#### 5.1.2. *Cross-site incompatibility*

If a user publishes personal information in multiple OSNs, she may employ different privacy control rules provided by different OSNs. This causes an inherent exploit named *cross-site incompatibility*.

**Exploit 2**. *Personal information could be inferred in multiple OSNs if it is protected by incompatible privacy rules in different OSNs.*

The incompatibility of privacy rules in different OSNs is due to: 1) inconsistent privacy rules in different OSNs, 2) different social relationships in different OSNs, and 3) different privacy control mechanisms in different OSNs (e.g. different privacy control granularities). Due to Exploit 2, an adversary may obtain a victim's personal particulars which are hidden from the adversary in one OSN but are shared with the adversary in another OSN. The following is an exemplary attack on Facebook and Google+.

**Attack 2**. Bob is Alice's friend on both Google+ and Facebook. On Google+, Alice publishes her gender information in her PP set and shares this information with some friends but not including Bob. On Facebook, Alice publishes her gender information and allows all users to view this information because Facebook allows her to share it with either all users or no users. Comparing Alice's personal information published on Facebook and Google+, Bob is able to know Alice's gender published on Facebook which is not supposed to be viewed by Bob on Google+.

Any adversary can perform this attack to infer personal information in a victim's *PP* set from multiple OSNs. This exploit can also be used to infer undisclosed information in *SR*

set and SA set. To prevent privacy leakage due to Exploit 2, the following necessary condition needs to be satisfied.

**Necessary Condition 2**. *Given a set of privacy rules* $PR = \{pr_1, pr_2, \ldots, pr_n\}$ *and* $pr_i = (wl_i, bl_i)$ *where* $pr_i$ *is the privacy rule for a victim's personal particular published in* $OSN_i$, $wl_i$ *is a set of all receivers in a white list, and* $bl_i$ *is a set of all receivers in a black list for* $i \in \{1,2,\ldots,n\}$, *the following condition holds: for any* $i,j \in \{1,2,\ldots,n\}$, $wl_i \setminus bl_i = wl_j \setminus bl_j$.[2]

To satisfy Necessary Condition 2, the following mitigation strategies can be applied.

**Mitigation 3**. *A victim should share her personal information with the same users in all OSNs.*

**Mitigation 4**. *If different OSNs provide incompatible privacy control on certain personal information, a victim should choose a privacy rule for this information under two requirements: 1) the privacy rule can be enforced in all OSNs; 2) the privacy rule is at least as rigid as the privacy rules which the victim intends to choose in any OSNs.*

### 5.2. SR set

A user's SR set records social relationships regarding whom the user knows. The undisclosed information in SR set protected by existing privacy control mechanisms can be inferred by two inherent exploits, namely *inferable social relationship* and *unregulated relationship recommendation*.

#### 5.2.1. *Inferable social relationship*

OSNs provide SR set for a user to store the lists of the users who have connections with him/her. If there exists a connection from Alice to Carl, then Carl is recorded in the outgoing list in Alice's SR set while Alice is recorded in the incoming list in Carl's SR set. The connection between Alice and Carl is stored in both Alice's SR set and Carl's SR set. This causes an inherent exploit named *inferable social relationship*, which corresponds to the information flow shown as dashed arrow (3) in Fig. 2.

**Exploit 3**. *Each social relationship in a victim's SR set indicates a connection between the victim and another user u. User u's SR set also stores a copy of this relationship for the same connection. The social relationship in the victim's SR set can be inferred from the SR set of another user who is in the victim's SR set.*

An adversary may use Exploit 3 to obtain undisclosed social relationships in a victim's SR set, which is shown in the following exemplary attack on Facebook.

**Attack 3**. Fig. 4 shows a scenario on Facebook, where Bob is a stranger to Alice, and Carl is Alice's friend. Alice shares her SR set with a user group including Carl. Bob guesses Carl may be connected with Alice, but cannot confirm this by viewing Alice's SR set as it is protected against him (who is a stranger to Alice). However, Carl shares his SR set to the public due to different concerns in privacy protection. Seeing Alice in Carl's SR set, Bob infers that Carl is Alice's friend.

Although the adversary is assumed to be a stranger in the above attack, any adversary with stronger capabilities can utilize Exploit 3 to perform the attack as long as he has two types of knowledge: 1) a list of users in the victim's SR set; 2) social relationships in these users' SR sets. This attack could be a stepping stone for an adversary to infiltrate a victim's social network. Once the adversary discovers a victim's friends and establishes connections with them, he becomes a friend of the victim's friends. After that, he has a higher probability to be accepted as the victim's friend, as they have common friends (Watts, 1999). To prevent privacy leakage caused by Exploit 3, the following necessary condition should be satisfied.

**Necessary Condition 3**. *Given a victim v's privacy rule* $pr_v = (wl_v, bl_v)$ *for her SR set, a set of all users* $U = \{u_1, u_2, \ldots, u_n\}$ *included in the victim's SR set in an OSN, and a set of privacy rules* $PR = \{pr_1, pr_2, \ldots, pr_n\}$ *where each* $pr_i = (wl_i, bl_i)$ *is the privacy rule for* $u_i$*'s SR set with white list* $wl_i$ *and black list* $bl_i$, *the following condition holds: for all* $i \in \{1,2,\ldots,n\}$, $wl_i \setminus bl_i \subseteq wl_v \setminus bl_v$.

To satisfy Necessary Condition 3, the following mitigation strategy can be applied.

**Mitigation 5**. *Let* $U = \{u_1, u_2, \ldots, u_m\}$ *denote the set of users in a victim's SR set. If the victim shares her SR set with a set of receivers, then each user* $u_i \in U$ *should share the social relationship between the user and the victim in the user's SR set with the same set of receivers only. Since most of existing OSNs use coarse-grained privacy rules to protect social relationships in SR set, all users in the victim's SR set should share their whole SR sets with the same set of receivers chosen by the victim in order to prevent privacy leakage.*

#### 5.2.2. *Unregulated relationship recommendation*

To help a user build more connections, most OSNs provide REC functionality to automatically recommend a list of other users whom this user may know. The recommendation list is usually calculated based on the relationships in SR set but not regulated by the privacy rules chosen by the users in the recommendation list. This causes an inherent exploit named *unregulated relationship recommendation*, which corresponds to the information flow shown as solid arrow (1) in Fig. 2.

**Exploit 4**. *All social relationships recorded in a victim's SR set could be automatically recommended by REC to all users in the victim's SR set, irrespective of whether or not the victim uses any privacy rules to protect her SR set.*

An adversary may use Exploit 4 to obtain undisclosed social relationships in a victim's SR set, which is shown in the following attack on Facebook.

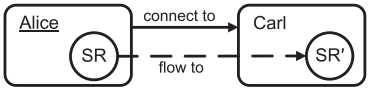


**Fig. 4 – Alice's social relationships flow to Carl's SR set.**

[2] Given a privacy rule $pr = \{wl, bl\}$ with a white list $wl$ and a black list $bl$, only the receivers who are in white list and are not in black list (i.e. any receiver $u \in wl \setminus bl$) are allowed to view the protected information.

**Attack 4**. On Facebook, Bob is a friend of Alice, but not in a user group named `Close_Friends`. Alice shares her SR set with `Close_Friends` only. Although Bob is not allowed to view Alice's social relationships in her SR set, such information is automatically recommended by REC to Bob as "*users he may know*". If Bob is connected with Alice only, the recommendation list consists of the social relationships in Alice's SR set only.

The recommendation list generated by *REC* may be affected by other factors such as personal particulars and interests, which may bring noise in social relationships. To minimize such noise, Bob could temporarily delete all his personal particulars and stay connected with the victim only.

The attack may happen on both Facebook and Google+ as long as an adversary is a *friend* of a victim. There is no prior knowledge required for this attack. The attack on Google+ is similar to the attack on Facebook but with a slight difference. On Facebook, the adversary cannot be connected with the victim unless the victim agrees since the relationship is mutual. By contrast, the adversary can set up a connection with the victim on Google+ without getting approval from the victim because the connection is unidirectional. This may make it easier for the adversary to obtain social relationships in the victim's *SR* set via *REC*.

We have reported Exploit 4 to Facebook and got confirmation from them. Exploit 4 occurs because *REC* functionality is implemented in a separate system not regulated by privacy control of Facebook. To prevent privacy leakage due to Exploit 4, the following necessary condition should be satisfied.

**Necessary Condition 4**. *Given a privacy rule $pr = (wl,bl)$ with white list $wl$ and black list $bl$ for a victim's SR set in an OSN and a set of all users $U$ included in the SR set, the following condition holds: $U \subseteq wl \setminus bl$.*

To satisfy Necessary Condition 4, the following mitigation strategy can be applied.

**Mitigation 6**. *Let $U = \{u_1,u_2,\ldots,u_m\}$ denote the set of users in a victim's SR set. If the victim shares her SR set with a set of users $U' \subseteq U$ only, the victim should remove any users in $U \setminus U'$ from her SR set in order to mitigate privacy leakage caused by REC.*

### 5.3. *SA set*

A user's *SA* set contains social activities about what the user does. The undisclosed information in *SA* set protected by existing privacy control mechanisms can be inferred due to the following inherent exploits and implementation defects, including *inferable social activity*, *ineffective rule update*, and *invalid hiding list*.

#### 5.3.1. *Inferable social activity*

If two users are connected in OSNs, a user's name can be mentioned by the other in a social activity via *TAG* such that this social activity provides a link to the profile page of the mentioned user. Such links create correlations among all the users involved in the same activity. This causes an inherent exploit named *inferable social activity*, which corresponds to the information flow shown as solid arrow (2) in Fig. 2.

**Exploit 5**. *If a victim's friend uses TAG to mention the victim in a social activity published by the victim's friend, it implies that the victim may also attend the activity, which is indicated by the link created by TAG pointing to the victim's profile page. Although this activity may involve the victim, the visibility of this activity is solely determined by the privacy rules specified by the victim's friend who publishes the activity, which is out of the control of the victim.*

An adversary may use Exploit 5 to obtain undisclosed social activities in a victim's *SA* set, which is shown in the following attack on Facebook.

**Attack 5**. Fig. 5 shows a scenario on Facebook, where Bob and Carl are Alice's friends, and Bob is Carl's friend. Alice publishes a social activity in her SA set regarding a party which Carl and she attended together and she allows Carl only to view this social activity. However, Carl publishes the same social activity in his SA set and mentions Alice via TAG. Due to different concerns in privacy protection, Carl allows all his friends to view this social activity. By viewing Carl's social activity, Bob can infer that Alice attended this party.

This attack works on Facebook, Google+, and Twitter. Any adversary can perform this attack if he knows the social activities published by the victim's friends pointing to the victim via *TAG*. To prevent privacy leakage due to Exploit 5, the following necessary condition should be satisfied.

**Necessary Condition 5**. *Given a privacy rule $pr_u = (wl_u,bl_u)$ for an activity where a victim $v$ is tagged by her friend $u$ in an OSN and $v$'s intended privacy rule $pr_v = (wl_v,bl_v)$ for the activity, the following condition holds: $wl_u \setminus bl_u \subseteq wl_v \setminus bl_v$.*

To satisfy Necessary Condition 5, the following mitigation strategy can be applied.

**Mitigation 7**. *If a victim is mentioned in a social activity in another user's SA via TAG, the victim should be able to specify additional privacy rules to address her privacy concerns even when the social activity is not in her profile page.*

#### 5.3.2. *Ineffective rule update*

It is common in OSNs that users regret sharing their social activities with wrong audience. Typical reasons include being in state of high emotion or under influence of alcohol (Wang et al., 2011). It is necessary to allow users to correct their

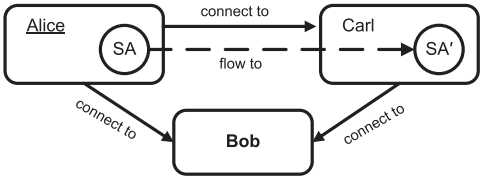


**Fig. 5 – Alice's social activities flow to Carl's *SA* set.**

mistakes by revoking the access rights of those unwanted audience. Once the access right of viewing a particular social activity is revoked, a receiver should not be able to view the activity protected by the updated privacy rule. On Facebook, a user can remove a receiver from the local white list specifying who is allowed to view a social activity or add the receiver to the local black list for the activity. Google+ and Twitter currently do not provide local black lists for individual social activities. A user may remove a receiver from the white list or from a user group if the user group is used to specify the scope of the white list (e.g. sharing a social activity within a *circle* on Google+). However, if a user's social activity has been pushed to her subscribers' feed pages, the update of privacy rules on Google+ and Twitter does not apply to this social activity in feed pages. This causes an implementation defect named *ineffective rule update*.

**Exploit 6**. *Once a victim publishes a social activity, the social activity is immediately pushed to the feed pages of the victim's subscribers who are allowed to view the social activity according to the victim's privacy rule. Later, even after the victim changes the privacy rule for this activity to disallow a subscriber to view this activity, the social activity still appears in this subscriber's feed pages on Google+ and Twitter. The current implementation of Google+ and Twitter enforces a privacy rule only when a social activity is published and pushed to corresponding subscribers' feed pages. Updated privacy rules are not applied to the activities which have already been pushed to feed pages (see Fig. 6).*

An adversary may use Exploit 6 to obtain undisclosed social activities in a victim's *SA* set without the victim's awareness. Below shows a typical attack on Google+.

**Attack 6**. On Google+, Bob is Alice's friend and subscriber. Alice publishes a social activity and allows her friends in group `Classmate` only to view the activity. Alice assigned Bob to the group `Classmate` by mistake and realized this mistake after publishing the activity. Then, Alice removed Bob from the group. However, Bob can still view this social activity as it has already been pushed to his feed page.

The above attack can happen on Google+ and Twitter. To perform the attack, an adversary should be the victim's *friend* and *subscriber*. The attack doesn't work on Facebook as privacy control in Facebook always actively examines whether privacy rule for a social activity is updated. If a privacy rule is updated, the privacy control is immediately applied to the social activity in corresponding feed pages.

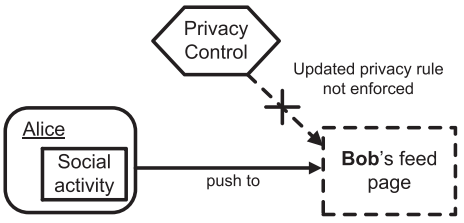


**Fig. 6 – Privacy control doesn't enforce the updated privacy rule to a social activity that has been pushed to a feed page.**

Consequently, the social activity is removed from the feed pages. To prevent this attack in certain OSNs such as Google+ and Twitter, the following mitigation strategy can be applied.

**Mitigation 8**. *If a victim mistakenly shares a social activity with an unintended receiver, instead of changing the privacy rules, the victim should delete the social activity as soon as possible so that the social activity is removed from all feed pages.*

Note that Mitigation 8 is not effective unless the deletion of the social activity takes place before an adversary views the social activity. If the adversary views the social activity before it is deleted, the adversary could keep a copy of this activity, which cannot be prevented.

#### 5.3.3. *Invalid hiding list*

To support flexible privacy control, many OSNs enable users to use black lists so as to hide information from specific receivers. On Facebook, a local black list is called *hiding list*. Using hiding list, a user may apply fine-grained privacy control on various types of personal information. However, the hiding lists take no effect except for the user's friends. This causes an implementation defect named *invalid hiding list*.

**Exploit 7**. *In certain OSN, a victim may include some of her friends in hiding lists to protect her personal information. However, when a friend breaks his relationship with the victim, the OSN automatically removes him from the hiding lists as the friend relationship terminates. Releasing from hiding lists, this former friend is allowed to view the victim's protected information if he is not restricted by other privacy rules.*

The implementation defect behind this exploit creates a false impression on the effectiveness of hiding lists. An adversary may use Exploit 7 to obtain undisclosed social activities in a victim's *SA* set without the victim's awareness. A typical attack on Facebook is given below.

**Attack 7**. On Facebook, Bob and Carl are Alice's friends. Bob is Carl's friend, which means Bob is also a friend of Alice's friend. Alice publishes a social activity which allows her friends and her friends-of-friends to view, except that Bob is added to the hiding list of this activity. Although Bob cannot view this activity under the current privacy rule, he can break his connection with Alice. Then, he is automatically removed from the hiding list. After that, Bob is able to view the undisclosed activity since he is a friend of Alice's friend.

Note that this attack does not work on Google+ and Twitter because their current privacy control mechanisms do not support any local black lists. Also note Exploit 7 can be exploited to target at not only *SA* set, but also *PP* set and *SR* set.

We have reported Exploit 7 to Facebook and received a confirmation from them.[3] To prevent this attack in affected OSNs such as Facebook, the following mitigation strategy can be applied.

[3] Exploit 7 has been fixed by Facebook in 2013.

**Mitigation 9**. *A victim should avoid using hiding lists when protecting personal information. Instead, a victim may use white lists or global black lists in forming privacy rules.*

## 6. Analysis of potentially vulnerable users

A user's personal information in OSNs could be leaked to adversaries who acquire necessary capabilities to perform the attacks, which have been discussed in Section 5. The effectiveness of the attacks can be affected by users' and their friends' sharing behaviors in OSNs. To investigate the users who can be vulnerable to these attacks, we conducted an online survey and collected users' usage data on Facebook, Google+, and Twitter. In this section, we first describe the design of the online survey. We then present the demographic data collected in the survey. Based on the survey results, we analyze how widely the users in OSNs can be vulnerable to the corresponding attacks.

### 6.1. Methodology

The participants to our online survey are mainly recruited from undergraduate students in our university. We mainly focus on young students in our survey because they are active users of OSNs. Our study shows that they are particularly vulnerable to the privacy attacks. Each participant uses at least one OSN among Facebook, Google+, and Twitter.

The survey questionnaire consists of four sections including 37 questions in total. In the first section, we gave an initial set of demographic questions and a set of general questions such as participants' awareness on privacy and what OSNs (i.e. Facebook, Google+, and Twitter) they use. All the participants need to answer the questions in the first section. In the following three sections, questions about participants' knowledge and privacy attitude towards Facebook, Google+, and Twitter are raised, respectively. Each participant only needs to answer the questions which are relevant to them in these three sections.

### 6.2. Demographics

There are 97 participants in total, among which 60 participants reported being male, and 37 reported female. Our participants' age ranges from 18 to 31, with an average of 22.7.

All of the 97 participants are Facebook users, among whom 95 participants have been using Facebook for more than 1 year, and 2 have been using Facebook for less than 1 month. About a half participants (41/97) are Google+ users, among whom 23 participants have been using Google+ for more than 1 year, 13 have been using Google+ for about 1 month−1 year, and 5 have been using Google+ for less than 1 month. Similarly, about a half participants (40/97) are Twitter users, among whom 36 participants have been using Twitter for more than 1 year, 3 have been using Twitter for about 1 month−1 year, and 1 has been using Twitter for less than 1 month.

### 6.3. Attacks to PP set

To obtain the undisclosed personal information in a victim's *PP* set, adversaries could exploit the inferable personal particulars and cross-site incompatibility to launch two corresponding attacks as discussed below.

#### 6.3.1. Inferable personal particulars

As discussed in Section 5.1.1, due to inferable personal particular (Exploit 1), a victim and most of his/her friends may share common or similar personal particulars. Our study results show that 71% of the Facebook users are connected with their classmates on Facebook; 78% of the Google+ users are connected with their classmates on Google+; and 73% of the Twitter users are connected with their classmates on Twitter.

Via Exploit 1, an adversary could perform Attack 1 and infer a victim's personal particular from the personal particulars shared by most of her friends. To perform Attack 1, two types of knowledge are required: a large portion of users stored in the victim's SR set and their personal particulars.

The protection of the victim's SR set could help prevent the adversary from obtaining the victim's relationships. Unfortunately, our study shows that 22% of the Facebook users, 39% of the Google+ users, and 35% of the Twitter users choose the "*Public*" privacy rule or the default privacy rule[4] for their social relationships, which means that these users share their social relationships with the public. Moreover, the OSNs users may connect to strangers. According to our study, 60% of the Facebook users, 27% of the Google+ users, and 30% of the Twitter users have set up connections with strangers, which leave their SR set information vulnerable to Exploit 4 (unregulated relationship recommendation) as discussed in Section 5.2.2.

The privacy rules for personal particulars of the victim's friends can be set to prevent the adversary from obtaining the second type of knowledge required in Attack 1. However, the victim's personal particulars can be exposed to threats if his/her friends publicly share their personal particulars. In our study, 43% of the Facebook users, 44% of the Google+ users, and 48% of the Twitter users share their personal particular publicly because they choose the "*Public*" privacy rule or the default privacy rule.[5]

#### 6.3.2. Cross-site incompatibility

Users may use multiple OSNs at the same time. According to our survey, 54 out of 97 participants use at least two OSNs as shown in Fig. 7. And 27 participants publish their posts in more than one OSN at the same time as shown in Fig. 8. If a user publishes personal information in multiple OSNs, he/she may set different privacy control rules vulnerable to Exploit 2, i.e. cross-site incompatibility.

[4] Facebook, Google+, and Twitter set "*Public*" as default privacy rule for the SR set of each user.

[5] Facebook, Google+, and Twitter set "*Public*" as the default privacy rule for each user's personal particulars such as "university" information.

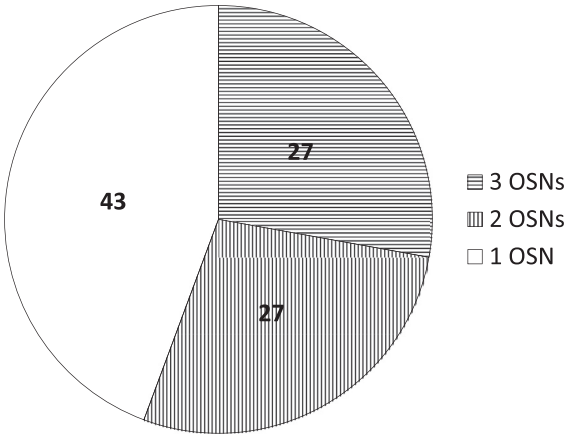


**Fig. 7 – Participants' usage of multiple OSNs.**

Due to Exploit 2, an adversary can perform Attack 2 if the victim shares her personal information with the adversary in any OSN site. This attack is due to three reasons.

The first reason is that users employ inconsistent privacy rules in different OSNs. The results of our study show that 27 out of 97 participants use inconsistent privacy rules to protect their gender information, 25 participants use inconsistent privacy rules to protect their university information, and 21 participants use inconsistent privacy rules to protect their political view information.

The second reason is that users maintain different social relationships in different OSNs. According to the study, 59 out of 97 participants reported that their social relationships on Facebook, Google+, and Twitter are different. Therefore, even though users protect their information by the same privacy rules on multiple OSNs, an adversary can still obtain their information if he can exploit this vulnerability.

The third reason is the difference between privacy control mechanisms in different OSNs. The protection of gender information is a typical example which is discussed in Section 5.1.2.

### 6.4. *Attacks to SR set*

Adversaries could obtain social relationships in a victim's *SR* set through two exploits, which are inferable social relationship and unregulated recommendation.

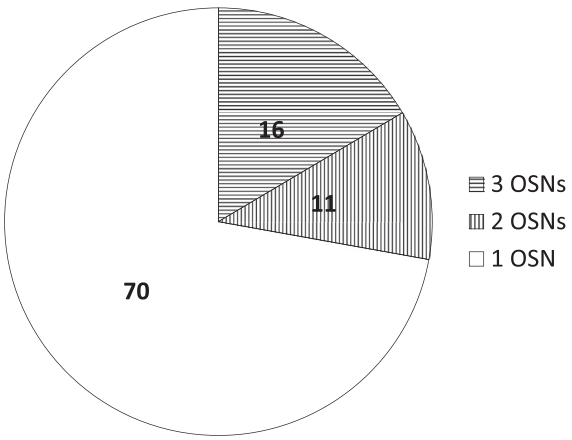


**Fig. 8 – Participants' publishing posts in multiple OSNs.**

#### 6.4.1. *Inferable social relationship*

Inferable social relationship (Exploit 3) is caused by the storage format of social relationships in *SR* set as explained in Section 5.2.1. If two users set up a relationship with each other, then each of them stores a copy of the relationship in his/her *SR* set and choose a privacy rule to protect his/her *SR* set.

Via Exploit 3, an adversary could perform Attack 3 given two types of knowledge, including a list of users in the victim's *SR* set and the social relationships in these users' *SR* set. Therefore, the protection of the social relationships in the victim's *SR* set depends on the privacy rules for the *SR* sets of the users in the victim's *SR* set. Unfortunately, as mentioned in Section 6.3.1, 22% of the Facebook users, 39% of the Google+ users, and 35% of the Twitters share their *SR* sets publicly. These users reveal social relationships with their friends publicly regardless of the privacy rules for their friends' *SR* sets.

#### 6.4.2. *Unregulated relationship recommendation*

*REC* functionality helps users establish more social relationships. According to our study, 71 out 97 Facebook users, 21 out of 41 Google+ users, and 17 out of 40 Twitter users have used *REC* functionality in OSNs. Unregulated relationship recommendation (Exploit 4) could leak all social relationships in a user's *SR* set due to automatic relationship recommendation of *REC*.

By Exploit 4, an adversary can perform Attack 4 to obtain all social relationships in a victim's *SR* set on Facebook or Google+ if the adversary manages to become a "friend" of the victim.

As shown in Fig. 9, 4% of the Facebook users and 7% of the Google+ users choose to share their *SR* set with a proper subset of their friends.[6] Exploit 4 explicitly violates these users' privacy rules.

Although most of the Facebook users and the Google+ users share their *SR* sets with friends, friends of friends, or public, their selection of privacy rules may contradict their privacy attitude.

In Fig. 9, 53% of the Facebook users share their *SR* sets with friends of friends or publicly.[7] Among the Facebook users who share their *SR* sets with friends of friends or public, 88% of them address concerns about their social relationships being revealed to others whom they don't know.

Among the Google+ users, 36% of them share their *SR* sets with friends of friends or the public. However, 71% of the Google+ users who share their *SR* sets with friends of friends or the public are not willing to reveal their social relationships to strangers.

As shown in our survey, 43% of the Facebook users and 20% of the Google+ users have concerns about revealing their social relationships to strangers but ever including strangers to their *SR* sets. This may leak the users' social relationships to the strangers irrespective of any privacy rules chosen to protect their *SR* sets.

[6] An empty subset corresponds to the privacy rule "*Only me*".

[7] The "*Public*" privacy rule and the default privacy rule lead to sharing *SR* set publicly.

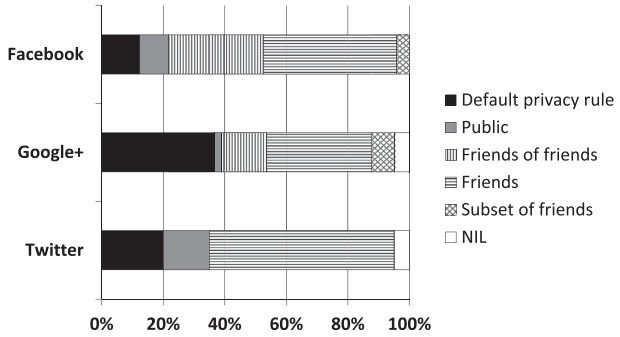


**Fig. 9 – Privacy rules for participants' SR sets in OSNs.**

### 6.5. *Attacks to SA set*

To obtain social activity information in a victim's SA set, adversaries could perform 3 attacks due to three exploits including inferable social activity, ineffective rule update, and invalid hiding list.

#### 6.5.1. *Inferable social activity*

In OSNs, if a user is mentioned in his/her friends' social activity via TAG, the privacy rule for the activity is determined by the friends and out of this user's control. This leads to inferable social activity (Exploit 5).

Via Exploit 5, an adversary may infer a victim's social activities from the victim's friends' SA set. As shown in Fig. 10, 99% of the Facebook users, 44% of the Google+ users, and 78% of the Twitter users have experience of being tagged in activities. On the other hand, 36% of the Facebook users, 34% of the Google+ users, and 40% of the Twitter users have concerns about being tagged in certain activities published by their friends without any negotiations. Since their friends determine the visibility of the activities, these users can inform their friends of their concerns. Our results show that 82% of the Facebook users, 73% of the Google+ users, and 73% of the Twitter users will inform their friends of their concerns if they don't agree on being tagged by their friends. The rest of them keep silent even though their privacy could be violated.

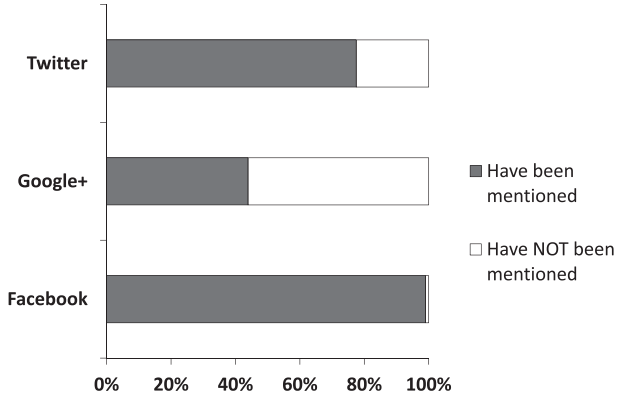


**Fig. 10 – Participants being mentioned in OSNs.**

#### 6.5.2. *Ineffective rule update*

As discussed in Section 5.3.2, if a user changes his/her privacy rules for social activities, the updated privacy rules do not apply to the activities which have been pushed to the feed pages of the user's subscribers. This is named as ineffective rule update (Exploit 6).

Via Exploit 6, an adversary could perform Attack 6 on Google+ and Twitter and obtain a victim's activities which are shared with the adversary before privacy rules update.

Changing privacy rules may occur if users regret publishing their activities. According to our study, 15% of Google+ users and 15% of Twitter users have experience of regretting publishing their posts. As shown in Fig. 11, 20% of the Google+ users choose to change their privacy rules if they regret sharing activities, while 38% of the Twitter users choose to change their privacy rules by turning on the protect my tweets option if they regret sharing such activities.

To mitigate Exploit 6, users may delete the activities they regret sharing as soon as possible. We found that 61% of the Google+ users and 23% of the Twitter users choose to do so.

#### 6.5.3. *Invalid hiding list*

On Facebook, if a user protects his/her social activity by using a hiding list including the user's friends, these friends will be automatically removed from the hiding list after they terminate their relationships with the user. This is referred to as the invalid hiding list (Exploit 7).

Via Exploit 7, an adversary could perform Attack 7 to obtain a victim's social activities if the victim uses the "*friends of friends*" privacy rule with a hiding list containing the adversary. Our study shows that 54% of the Facebook users have ever used the "*friends of friends*" privacy rule with a hiding list that includes their friends when they publish activities. To evaluate the awareness of the risks caused by using the invalid hiding list, we summarized participants' confidence level regarding whether their activities are hidden from their friends who are included in their hiding lists on Facebook. As shown in Fig. 12, 31% (30 out of 97) of the Facebook users feel confident in the effectiveness of the hiding list on Facebook. If attack 7 happens, these participants may misunderstand the

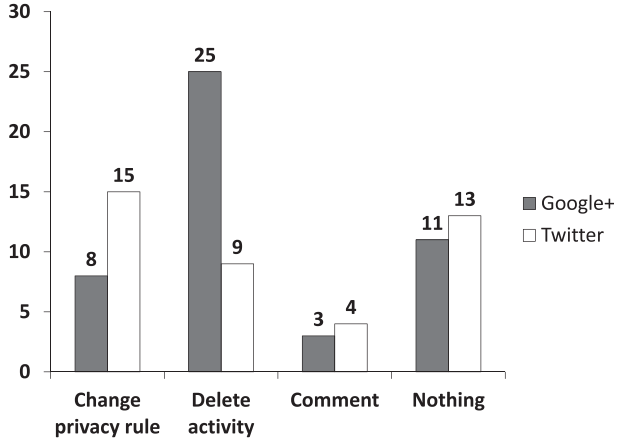


**Fig. 11 – Participants' actions if regretting sharing activities.**

validity of the hiding lists and still believe that their activities are hidden from their friends included in the hiding lists.

## 7. Discussion

On the surface, our exploits are caused by the inconsistencies between privacy control and functionalities of OSN. In fact, these inconsistencies reflect the conflicts between users' intention on privacy protection and social/business values of OSNs. We discuss the implications of these conflicts and the impacts on users' sharing behaviors due to the conflicts in this section.

Most of the functionalities involved in our exploits are essential in OSNs. These functionalities deal with personal particulars, social relationships, and social activities. While the social values of these functionalities should be preserved from a user's perspective, they are restricted due to privacy controls.

First, exhibiting personal particulars is an important feature for social recognition. Most OSNs encourage users to share genuine information about their personal particulars in order to foster trust and respect in OSNs (FaceBook and http://sec.gov/, 1326). This would help users discover new relationships with those who have similar interests. This is explained by the homophily theory (McPherson et al., 2001; Centola et al., 2007), which states that a human being is more willing to interact with others who have similar personal particulars such as race, organization, and education. Meanwhile, the implicit connections among users may be exploited to infer undisclosed personal particulars. According to Yamada et al. (2012); Zheleva and Getoor (2009); Dey et al. (2012), 62% of users consider that their personal particulars published in OSNs are sensitive. To mitigate this threat for these users, **mitigations** 1 and 2 require them to connect with other dissimilar users which they may not even like.

Second, maintaining and expanding social relationships is one of the major benefits of OSNs. As socially oriented beings, humans have a desire to stay connected so that they have a sense of communion with others (Sheldon and Bettencourt, 2002). This desire is addressed in OSNs with the relationship list and the recommendation function. Although the public display of a user's relationship list may disclose certain private information, it also helps build more connections in OSNs. If a user's profile contains a large number of connections, it brings satisfactory social recognition for the user (Hei-man, 2008). The recommendation function further makes it easier to establish new connections based on relationship lists and other information. This is especially important for new users to make friends in OSNs. The current recommendation function operates according to the small-world theory (Watts, 1999), which states that two connected users are likely to have common friends who have not yet recorded in their current relationship lists. This function can also be exploited by an adversary to enumerate all social relationships of a victim. 45% of users believe that their social relationships in OSNs are sensitive (Yamada et al., 2012; Dey et al., 2012). Since the disclosure of the social relationships can be a stepping stone for advance attacks on personal particulars and social activities, the protection of the social relationships is also important to those who consider that personal particulars and social activities are sensitive (Yamada et al., 2012; Dey et al., 2012; Zheleva and Getoor, 2009). To mitigate the privacy leakage about social relationships, a user may use **mitigations** 5 and 6. The consequences of applying these mitigation strategies are: 1) If a user sets up a strict privacy rule on his relationship list, this rule should propagate to all users in his relationship list. 2) The effectiveness of the recommendation function would be significantly influenced by such mitigations.

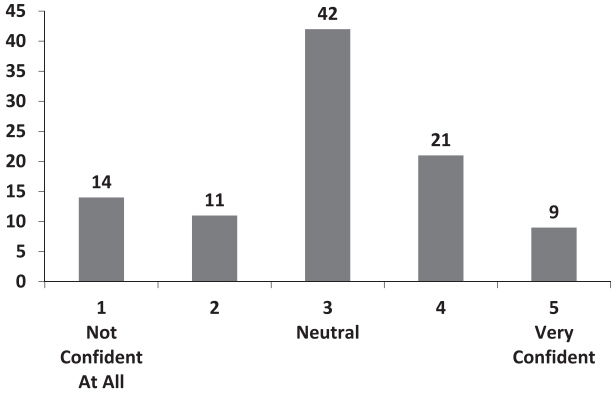


**Fig. 12 – Users' confidence in validity of Facebook hiding list.**

Third, sharing social activities is an important part of human social life. Human beings are curious about what happen around them. They would like to understand the surrounding environment by knowing how other people behave, think, and feel (Renner, 2006). OSNs enable users to receive the activities published by other users to cure such curiosity. On the other hand, users who publish activities feel rewarded due to attentions of other users, which is usually interpreted as a sign for social recognition (Hotz). Since a social activity usually involves multiple users, sharing this activity may conflict with the privacy concern of these users. The social activities are considered as sensitive information by 66% of users (Yamada et al., 2012; Dey et al., 2012). For these users, in order to mitigate this threat, the scope of privacy control in OSNs should be extended as mentioned in **mitigation** 7, which enforces privacy control to an activity no matter who publishes it. An application of **mitigation** 7 can be privacy policy negotiation mechanisms (Wishart et al., 2010; Hu et al., 2011) which seeks tradeoff between the users' privacy intention and their desire to share social activity by requiring the users to choose their privacy rules and sensitivity of each activity before publishing the activity. However, this may frustrate users who intend to share that activity, and might be difficult to achieve due to the incompatibility among privacy control mechanisms in different OSNs. As suggested in **mitigations** 3 and 4, a user may choose a strict privacy rule so as to achieve his privacy objective. However, this may significantly restrict the sharing nature of OSNs.

While OSN users are concerned with the social values of OSN functionalities, OSN service providers are more concerned with business values. As a company, the first priority of an OSN service provider is to generate revenue. However, most existing OSN service providers do not charge their

users. As Andrew Lewis pointed out, "If you're not paying for something, you're not the customer; you're the product being sold." This is exactly what OSN service providers do, monetizing user-generated contents by maintaining an OSN-based ecosystem. One of the most successful OSN-based business models, targeted advertising (Spaulding, 2010), usually demands high quality of personal information and large number of connected individuals (FaceBook and http://sec.gov/, 1326; Brown, 1950; Coffin, 1963; Wolin and KorgaonkarBhat, 2003; Danaher and. Mullarkey, 2003). Thus, an OSN service provider has strong incentive to encourage users to share personal information, and connect to more users.

To address users' concern about disclosure of sensitive personal information, the existing OSN service provider, such Facebook and Google+, provide fine-grained privacy rules for users. As these privacy rules become more complex, users could leak their sensitive information due to misunderstanding and misusing the privacy rules (Wang et al., 2011). Thus efforts are spent on improving the usability of the privacy control by automatic content-based reminders (Wang et al., 2011; Sinha et al., 2013). For example, an automatic privacy rule recommendation tool is proposed to deduce users' sharing preferences based on the users' sharing content in order to help users avoid misconfiguration of the privacy rules (Sinha et al., 2013). However, the above methods cannot completely resolve the identified exploits. And these methods and almost all mitigations discussed in the paper add additional restrictions on user generated contents published or shared in OSNs.

## 8. Related work

Due to wide adoption of OSNs, the privacy problem of OSNs has attracted strong interest among researchers. We summarize the related work in this area in terms of attacks, privacy settings, and access control models.

The attack techniques proposed in prior literature mainly focus on inferring users' identity (Backstrom et al., 2007) and other personal information (Zheleva and Getoor, 2009; Balduzzi et al., 2010; Chaabane et al., 2012) from public information shared in OSNs. Zheleva and Getoor (2009) proposed a classification-based approach to infer users' undisclosed personal particulars from their social relationships and group information which are publicly shared. Chaabane et al. (2012) proposed to infer users' undisclosed personal particulars from public shared interests and public personal particulars of other users who have similar interests. Balduzzi et al. (2010) utilized email addresses as unique identifiers to identify and link user profiles across several popular OSNs. Since users' information may be shared publicly in an OSN but not be shared in another OSN, certain hidden information can be revealed by combining public information collected from different OSNs. The effectiveness of these attacks largely depends on the quality of public information, which can be affected due to users' awareness of privacy concerns. As reported in Chaabane et al. (2012), only 18% of Facebook users now publicly share their social relationships and 2% of Facebook users publicly share their dates of birth. Thus it is more realistic to analyze the threats caused by more powerful adversaries or insiders as in our analysis.

The threat of privacy leakage caused by insiders is also mentioned by Johnson et al. (2012). They investigated users' privacy concerns on Facebook and discovered that the privacy control mechanisms in existing OSNs help users manage outsider threats effectively but cannot mitigate insider threats because users often wrongly include inappropriate audiences as members of their friend network. Wang et al. (2011) analyzed reasons why users wrongly configure privacy settings and provided suggestions for users to avoid such mistakes. To help users handle complex privacy policy management, Cheek and Shehab, (2012) proposed two approaches using clustering techniques to assist users in grouping friends and setting appropriate privacy rules. However, as shown in our work, privacy leakage could still happen even if a user correctly configures his privacy settings due to the exploits caused by inherent conflicts between privacy control and OSN functionalities.

Some researchers addressed the privacy control problem in traditional access control modeling. Several models (Fong et al., 2009; Carminati et al., 2009) are established to provide more flexible and fine-grained control so as to increase the expressive power of privacy control models. Nevertheless, this is not sufficient to guarantee effective privacy protection. From our analysis on information flows, OSN functionalities may be affected by privacy control. On the other hand, a more complex privacy control model increases users' burden on configuring privacy rules.

One of the exploits found in our work (Exploit 5) is also mentioned in previous research on resolving privacy conflicts in collaborative data sharing. Wishart et al. (2010) and Hu et al. (2011) analyzed co-owned information disclosure due to conflicts of privacy rules set by multiple owners. They also introduced a negotiation mechanism to seek a balance between the risk of privacy leakage and the benefit of data sharing. Compared to them, our work investigates a broader range of privacy threats in OSNs, discovers the underlying conflicts between privacy control and social/business values of OSNs, and analyzes the difficulty in resolving these conflicts, which have not been addressed in previous works.

## 9. Conclusion

In this paper, we investigated privacy leakage under privacy control in online social networks. Our analysis showed that privacy leakage could still happen even after users correctly configure their privacy settings. We examined real-world OSNs including Facebook, Google+, and Twitter, and discovered the exploits which lead to privacy leakage. Based on the findings, a series of attacks were introduced for adversaries with different capabilities to learn undisclosed personal information. We analyzed necessary conditions and provided suggestions for users to mitigate privacy leakage in OSNs. We conducted a user study to investigate the potentially vulnerable users due to the attacks. In the end, we discussed the implications of resolving privacy leakage in OSNs.

## Appendix A. Proof of Necessary Condition 1

The input of an adversary includes two types of knowledge about a victim: a subset $U = \{u_1,u_2,\ldots,u_n\}$ of a victim $v$'s SR set in an OSN, and personal particular value $pp_{u_i}$ ($pp_{u_i} \neq null$) of each receiver $u_i \in U$. The adversary may infer the victim's personal particular $pp_v$ ($pp_v \neq null$) by calculating the common personal particular value shared by most of the victim's friends with Algorithm 1.

**Algorithm 1** Infer Personal Particular

**Input:** $U = \{u_1, u_2, ..., u_n\}$; $pp_{u_1}, pp_{u_2}, ..., pp_{u_n}$;
**Output:** $pp_{infer}$

1: compute $PP = \{pp_1, pp_2, ..., pp_m\}$ from $pp_{u_i}$ for all $i \in \{1, 2, ..., n\}$
2: **for all** $j \in \{1, 2, ..., m\}$ **do**
3: calculate $U_{pp_j} \subseteq U$ such that for all $u \in U_{pp_j}$, $pp_u = pp_j$
4: **end for**
5: **if** there exists $U_{pp_t}$ such that $|U_{pp_t}| > |U_{pp_s}|$ for all $s \in \{1, 2, ..., m\}$ and $t \neq s$ **then**
6: **return** personal particular value $pp_t$
7: **else**
8: **return** null
9: **end if**

Given the inputs, if Algorithm 1 returns a value $pp_{infer}$ which is equal to the victim's personal particular $pp_v$, then the victim's personal particular information is leaked to the adversary.

## Appendix B. Proof of Necessary Condition 2

A victim uses the privacy rules $pr_1,pr_2,\ldots,pr_n$ to protect her personal particular published in $OSN_1,OSN_2,\ldots,OSN_n$, respectively where each privacy rule $pr_i = (wl_i,bl_i)$ contains a white list $wl_i$ and a black list $bl_i$. Assuming there are two privacy rules $pr_t$ and $pr_j$ such that $(wl_t \setminus bl_t \neq wl_j \setminus bl_j)$ where $t,j \in \{1,2,\ldots,n\}$ and $t \neq j$, we have $U_{diff} = (wl_t \setminus bl_t) \setminus (wl_j \setminus bl_j) \neq \emptyset$. If an adversary $adv \in U_{diff}$, then the victim's personal information is leaked to the adversary although the information is supposed to be hidden from the adversary by $pr_j$ on $OSN_j$.

## Appendix C. Proof of Necessary Condition 3

A victim $v$ sets the privacy rule $pr_v = (wl_v,bl_v)$ for her SR set with white list $wl_v$ and black list $bl_v$. The victim's SR includes a set of users $U = \{u_1,u_2,\ldots,u_n\}$. Each user $u_i$ sets the privacy rule $pr_i = (wl_i,bl_i)$ for his/her SR set with white list $wl_i$ and black list $bl_i$ for all $i \in \{1,2,\ldots,n\}$. Assuming an adversary $adv$ is not in $wl_v \setminus bl_v$, the adversary is not allowed to view any relationships in the victim's SR set. If there is a privacy rule $pr_t$ such that $wl_t \setminus bl_t$ is not a subset of $wl_v \setminus bl_v$ and $t \in \{1,2,\ldots,n\}$, then we have $U_{diff} = (wl_t \setminus bl_t) \setminus (wl_v \setminus bl_v) \neq \emptyset$. Assuming $adv \in U_{diff}$, then the relationship between user $u_t$ and victim $v$ is known by adversary $adv$ although the information in the victim's SR set should be hidden from $adv$ by $pr_v$.

## Appendix D. Proof of Necessary Condition 4

A victim sets a privacy rule $pr_v = (wl_v,bl_v)$ for her SR set with white list $wl_v$ and black list $bl_v$. The victim's SR includes a set of users $U = \{u_1,u_2,\ldots,u_n\}$. Assuming that $U$ is not a subset of $wl_v \setminus bl_v$, then we have $U_{diff} = U \setminus (wl_v \setminus bl_v) \neq \emptyset$. If adversary $adv \in U_{diff}$, then REC functionality recommends almost all users in $U$ to $adv$. Note that these users should be hidden from $adv$ by privacy rule $pr_v$ because $adv$ is not in $wl_v \setminus bl_v$.

## Appendix E. Proof of Necessary Condition 5

Given a privacy rule $pr_u = (wl_u,bl_u)$ for an activity with white list $wl_u$ and black list $bl_u$ where victim $v$ is mentioned by her friend $u$, any receivers in $wl_u \setminus bl_u$ are allowed to view the activity. We assume that $v$'s intended privacy rule for the activity is $pr_v = (wl_v,bl_v)$ with white list $wl_v$ and black list $bl_v$. If $wl_u \setminus bl_u$ is not a subset of $wl_v \setminus bl_v$, then we have $U_{diff} = (wl_u \setminus bl_u) \setminus (wl_v \setminus bl_v) \neq \emptyset$. Assuming $adv \in U_{diff}$, then $adv$ can obtain the activity published by $u$ although the victim's privacy rule $pr_v$ prevents $adv$ from viewing the activity.

## Appendix A. Supplementary data

Supplementary data related to this article can be found at http://dx.doi.org/10.1016/j.cose.2014.10.012.

## REFERENCES


Aquino C, http://blog.comscore.com/2012/01/its _a_social_world.html.

Backstrom L, Dwork C, Kleinberg J. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In: Proceedings of the 16th international conference on World Wide Web; 2007. p. 181–90.

Balduzzi M, Platzer C, Holz T, Kirda E, Balzarotti D, Kruegel C. Abusing social networks for automated user profiling. In: Proceedings of the 13th international conference on recent advances in intrusion detection; 2010. p. 422–41.

Brown WF. The determination of factors influencing brand choice. J Mark 1950;14(5)):699–706.

Carminati B, Ferrari E, Heatherly R, Kantarcioglu M, Thuraisingham B. A semantic web based framework for social network access control. In: Proceedings of the 14th ACM symposium on access control models and technologies; 2009. p. 177–86.

Centola D, Gonzalez-Avella JC, Eguiluz VM, Miguel MS. Homophily, cultural drift, and the co-evolution of cultural groups. J Confl Resolut 2007;51(6):905–29.

Chaabane A, Acs G, Kaafar MA. You are what you like! information leakage through users interests. In: Proceedings of the 19th annual network & distributed system security symposium; 2012.

Cheek GP, Shehab M. Policy-by-example for online social networks. In: Proceedings of the 17th ACM symposium on access control models and technologies, SACMAT '12; 2012. p. 23–32.

Coffin TE. A pioneering experiment in assessing advertising effectiveness. J Mark 1963;27(3):1–10.

Danaher PJ, Mullarkey GW. Factors affecting online advertising recall: a study of students. J Advert Res 2003;43(3):252–67.
Dey R, Jelveh Z, Ross K. Facebook users have become much more private: a large-scale study. In: Proceedings of the 10th pervasive computing and communications workshops; 2012. p. 346–52.
FaceBook, http://sec.gov/Archives/edgar/data/1326801/000119312512034517/d287954ds1.htm.
Fong P, Anwar M, Zhao Z. A privacy preservation model for facebook-style social network systems. In: Computer Security – ESORICS 2009, vol. 5789; 2009. p. 303–20.
Hei-Man TSE. An ethnography of social network in cyberspace: the facebook phenomenon. Hong Kong Anthropol 2008;2:53–77.
Hotz RL, Science reveals why we brag so much, http://online.wsj.com/article/SB10001424052702304451104577390392329291890.html?mod=googlenews_wsj.
Hu H, Ahn G-J, Jorgensen J. Detecting and resolving privacy conflicts for collaborative data sharing in online social networks. In: Proceedings of the 27th annual computer security applications conference; 2011. p. 103–12.
Johnson M, Egelman S, Bellovin SM. Facebook and privacy: it's complicated. In: Proceedings of the eighth symposium on usable privacy and security; 2012. 9:1–9:15.
McPherson M, Smith-Lovin L, Cook JM. Birds of a feather: homophily in social networks. Annu Rev Sociol 2001;27:415–44.
Renner B. Curiosity about people: the development of a social curiosity measure in adults. J Personal Assess 2006;87(3):305–16.
Sheldon KM, Bettencourt BA. Psychological need-satisfaction and subjective well-being within social groups. Br J Soc Psychol 2002;41(1):25–38.
Sinha A, Li Y, Bauer L. What you want is not what you get: Predicting sharing policies for text-based content on facebook. In: Proceedings of the 2013 ACM workshop on artificial intelligence and security, AISec '13. ACM; 2013. p. 13–24.
Spaulding TJ. How can virtual communities create value for business? Electron Commer Res Appl 2010:38–49.
Tang C, Ross K, Saxena N, Chen R. Whats in a name: a study of names, gender inference, and gender behavior in facebook. In: Database Systems for Adanced Applications, Vol. 6637 of Lecture Notes in Computer Science. Springer Berlin Heidelberg; 2011. p. 344–56.
Wang Y, Norcie G, Komanduri S, Acquisti A, Leon PG, Cranor LF. "i regretted the minute i pressed share": a qualitative study of regrets on facebook. Proc Seventh Symp Usable Priv Secur 2011:10:1–10:16.
Watts DJ. Small worlds: the dynamics of networks between order and randomness. Princeton University Press; 1999.
Wishart R, Corapi D, Marinovic S, Sloman M. Collaborative privacy policy authoring in a social networking context. In: Proceedings of the 2010 IEEE international symposium on policies for distributed systems and networks; 2010. p. 1–8.
Wolin LD, KorgaonkarBhat P. Web advertising: gender differences in beliefs, attitudes and behavior. Internet Res 2003;13(5):375–85.
Yamada A, Kim TH-J, Perrig A. Exploiting privacy policy conflicts in online social networks. In: Technical report. Carnegie Mellon University; 2012. p. 1–9.
Zheleva E, Getoor L. To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles. In: Proceedings of the 18th international conference on world wide web; 2009. p. 531–40.

**Yan Li** is currently a Research Fellow at Singapore Management University. He received his Ph.D. degree in Information Systems from Singapore Management University in July 2014. His current research interests include privacy and security in social networks, face biometric-based user authentication, and liveness detection for face authentication.

**Yingjiu Li** is currently an Associate Professor in the School of Information Systems at Singapore Management University. He received his Ph.D. degree in Information Technology from George Mason University in 2003. His research interests include RFID security, applied cryptography, and data applications security. He has published over 70 technical papers in international conferences and journals. He has served in the program committees for over 50 international conferences and workshops. Yingjiu Li is a senior member of the ACM and a member of the IEEE.

**Qiang Yan** is currently a privacy engineer in Google. He received his Ph.D. degree in Information Systems from Singapore Management University in 2013. His research interests include mobile security and privacy, human factors insecurity system design, and security and privacy issues in social networks.

**Robert H. Deng** is currently a professor, associate dean for Faculty and Research, School of Information Systems at Singapore Management University. He received his Ph.D. degrees from the Illinois Institute of Technology. He has more than 200 technical publications in international conferences and journals in the areas of computer networks, network security, and information security. He has served as general chair, program committee chair, and program committee member of numerous international conferences. He is an Associate Editor of the IEEE Transactions on Dependable and Secure Computing, Associate Editor of Security and Communication Networks Journal (John Wiley).