My Friend Leaks My Privacy: Modeling and Analyzing Privacy in Social Networks

Manuscript submitted for double-blind review.

ABSTRACT

With the dramatically increasing participation in online social networks, huge amount of private information becomes available on such sites. It is critical to preserve users' privacy, without preventing them from socialization and information sharing. Unfortunately, existing approaches fall short meeting such requirements.

We argue that the key component of privacy protection in social networks is protecting (sensitive) content, i.e. privacy as having the ability to control dissemination of information. We follow the concepts of *private information boundaries* and *restricted access and limited control* to introduce a *social circle* model. We articulate the formal constructs of this model, and enumerate the desired properties for privacy protection in the model. We show that the social circle model is efficient yet practical, which provides certain level of privacy protection capabilities to social network users, while still facilitates socialization. We then utilize this model to analyze the most popular social network platforms on the Internet (Facebook, Google+, VK, WeChat, Sina Weibo, etc), and demonstrate the potential privacy vulnerabilities in some social networks. Finally, we discuss the implications of the analysis, and possible future directions.

ACM Reference format:

Manuscript submitted for double-blind review.. 2016. My Friend Leaks My Privacy: Modeling and Analyzing Privacy in Social Networks. In *Proceedings of ACM Conference, Washington, DC, USA, July 2017 (Conference'17),* 12 pages.

DOI: 10.1145/nnnnnnn.nnnnnnn

1 INTRODUCTION

In recent years, many online social network services (SNSs) such as Facebook have become extremely popular, attracting a large number of users to generate and share diverse (personal) contents. With the advancement of information retrieval and search techniques, on the other hand, it has become easier to do web-scale identification and extraction of users' personal information from SNSs. Therefore, malicious or curious users could take an advantage of these techniques to collect others' private and sensitive information. In fact, we have been overwhelmed by news reports on the problems caused by the lack of social network privacy. Let us illustrate a real-world case where one's private information is being leaked.

Example 1.1. **(Private Information Disclosure)** As shown in Fig. 1 (a), the owner of information, say Alice, shares two photos on Google+. She is cautious about her privacy so that she configures the album to be available only to a limited audience (i.e., small circle of friends), which includes Mallory. However, in Google+ and other SNSs, friends are often allowed to re-share their friends' photos, which potentially redefines the privacy setting set by the original content owner. In this example, although Mallory receives a warning when she attempts to re-share Alice's photo (Fig. 1 (b)), she can simply ignore the warning and re-share the photo with a different circle (Fig. 1 (c)). Now, Chuck, who is not a member of Alice's circle (so that he could not see Alice's original post), is able to see the photos from Mallory's wall (Fig. 1 (d)). Worst of all, if Alice is not a member of Mallory's circle, she does not get notified of the re-share (Fig. 1 (e)). Although, it is possible for Alice to disable re-share, that function is not obvious to regular users and it needs to be explicitly invoked for each post, which significantly degrades the level of usability.

As we have demonstrated, for various design rationales and business decisions, some SNSs promote information sharing aggressively, to the extend that introduces privacy vulnerabilities. Similar privacy breach has existed in Facebook until 2014, without the warning message or the function to disable forwarding, ever since such functions were introduced [25, 57]. However, as we will demonstrate later, private information leakage is common in many SNSs including Google+ and Sina Weibo.

In the literature, other privacy protection models and mechanisms, such as k-anonymity [52] and differential privacy [13], have been developed for privacy-preserving data analysis. Such solutions protect individual user's identity information in statistical databases, so that adversaries cannot easily re-identify a user from sanitized datasets. However, they are not suitable to protect (sensitive) user contents in the settings of online social networks, where user IDs (or screen names) are revealed. Moreover, an ideal privacy protection solution for SNSs is *not* to discourage the socialization such as sharing photos with friends. In this context, behavioral researchers and practitioners argue that privacy could be defined as *having the ability to control the dissemination of (personal) information.* Recently, the concept of social circles have been adopted in the research community [39, 48, 49] and in commercial products [26, 55]. The key idea is that new messages are posted to designated audience (i.e., social circles) and the message owners have a full control of the information boundary, where information is conceptually bounded by the social circle. Meanwhile, social circles are also expected to promote information sharing, since they give users the perception of security and privacy. However, social circles are neither clearly defined nor strictly enforced (e.g., circle leakage in Example 1.1). [55] also indicate that use of social circles is limited due to lack of users, and users are unaware of how information could spread beyond circles in Google+. We argue that current adoptions of social circles have significant drawbacks: (1) the social circle model was loosely defined and there was no formal underpinning to support the model; (2) There was no systematic analysis of the requirements, properties, and issues associated with the model; (3) There is a major usability issue that prevents users from adopting social circles: it is labor-intensive and tedious to manually arrange existing users into circles, and to identify the appropriate circle for

Conference'17, Washington, DC, USA
2016. 978-x-xxxx-xxxx-x/YY/MM…$15.00
DOI: 10.1145/nnnnnnn.nnnnnnn

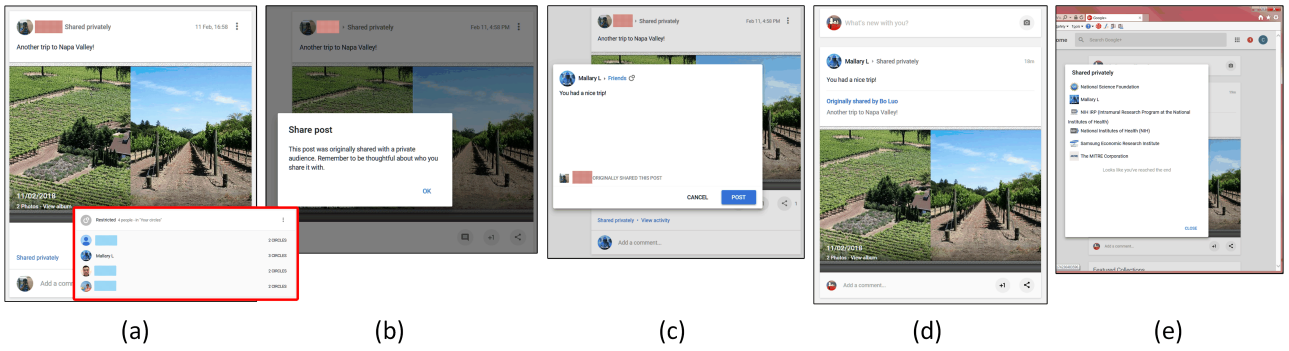

Figure 1: The demonstration of privacy breach on Google+ (Example was captured on 02/01/2018): (a) information owner (Alice) posts photos to a circle of 4 users, including Mallory; (b) Mallory attempts to re-share and sees a warning; (c) Mallory ignores the warning and re-shares the photos; (d) Chuck, who cannot see Alice's post, now sees the photos from Mallory; (e) The photos are re-shared to a completely different circle, which *excludes* Alice.

every new message; and (4) There is no available solution to detect leaky circles (as in Example 1.1) – users are more vulnerable since their perceived protection boundaries are often quietly violated.

The contributions of this paper are three-fold: (1) we have formalized a *social circle* model for social network privacy protection, which is based on the notions of *private information boundaries* and *restricted access and limited control*. The social circle model facilitates socialization while allows users to control the dissemination of their information belongings; (2) Using the social circle model as the basis of analysis, we have carefully examined the privacy enforcement mechanisms of six leading social networking platforms; and (3) We further discuss the implications of our findings. The proposed model and analysis is expected to serve as a blueprint of technological approaches to improve the validity, usability and efficiency of social network privacy protection solutions.

2 THE SOCIAL CIRCLE MODEL

2.1 Preliminaries

Adoption of user privacy control presents not only a *technical* challenge, but also a *social* one. Studies have shown that even users with high concern about privacy do not always take appropriate actions even when those measures are fairly easy to perform. This phenomenon is known as the **Privacy Paradox**–i.e., users state high levels of privacy concerns but behave in ways that seemingly contradict their privacy attitudes [1, 2, 41]. Two complementary theoretical explanations have been proposed. First, Acquisti *et al.* argued that the dichotomy between privacy attitude and behavior is due to *bounded rationality*–i.e., human agents are unable to have absolute rationality because they either do not have the proper knowledge to process the risks, or they underestimate the risks by discounting the possibility of future negative events [1, 2]. Second, privacy control features make users' online profiles less visible, and thus can work against developing social relationships. This causes privacy control to be viewed as an additional cost in terms of social-relational concerns [14]. In either theoretical explanation, privacy control features will not be utilized if the costs are perceived to be greater than the benefits, despite users' privacy concerns. Therefore, an ideal privacy protection model that addresses the privacy paradox is expected to have sufficient rigor and expressiveness to satisfy the privacy expectations of the users, while it should also be easily understandable and highly usable.

There has been a movement toward the conceptualization of privacy as "the ability of individuals to control the terms under which their personal information is acquired and used" [10] (p.326). This concept of privacy-as-control originated in Westin's [56] and Altman's [3] theories of privacy, which have since entered the mainstream of privacy research in information systems, HCI, marketing, and sociology. The notion of privacy as control has, however, been criticized for its vagueness with regard to (1) the types of personal information over which people can expect to have control; and (2) the amount of control they can expect to have over their personal information [53]. These problems in defining privacy as control spurred the formulation of a modified notion of privacy as control and restricted access, which advocates for the provision of different levels of restricted access to different people for different types of information in different situations [53]. From this perspective, a privacy model known as *restricted access and limited control (RALC)* [53] has emerged. This new model highlights the need for the creation of a privacy protection boundary to enable people to restrict others from accessing their personal information [53].

2.2 The Social Circle Model

Drawing on the privacy theory of RALC, we argue that privacy is a multifaceted concept that should be analyzed with the considerations of: (1) degree of control over information dissemination; and (2) the extent to which their privacy (protection) expectations are met (perceived protection "boundaries"). The theoretical distinction between control and information boundary seems readily understood. However, most users in practice may conflate these two dimensions by having an "illusion" of control on the information they reveal: Because they have control over the information publication, they believe they also have control over the accessibility and use of that information by others. Such "illusion" could be explained by the optimistic bias where users overestimate their control over information dissemination, and meanwhile underestimate the future to their shared information by others. In addition, the optimistic bias may also be caused by the gap between users' perceived information boundary and the actual boundary enforced

by their privacy settings. The gap might be caused by: (1) the social network sites often adopt over-simplified privacy models, which fail to accurately capture users' perceptions; or (2) when more powerful privacy models are adopted, the actual implementations fail to correctly enforce the models.

The above social science theoretical perspectives reveal different but interrelated approaches to conceptualize a privacy model. When looking across these different aspects, we find that individual privacy is better viewed as a multifaceted concept with the considerations of: **(1)** the extent to which users can control over the disclosure, dissemination, and transitive propagation of their personal information (the strength of deployed privacy control mechanisms); **(2)** the extent to which their privacy (protection) expectations are met (perceived protection "boundaries"); and **(3)** the subjective estimation of the gap between users' perceptions of protection boundary and the actual boundary enforced by their privacy settings (optimistic bias).

Based on these findings, now, we formalize the model termed as **Social Circles** to integrate the *control perspective* and the *restricted access to information boundary perspective* as follows:

Definition 2.1 (Social Network Identity (SNID)). A social network identity (SNID) is defined as the identity a specific end user adopts for a given SNS (or social application). It consists of two attributes, the name of the SNS and the name the user has adopted for that site–e.g., SNID_1=(facebook, dul13);

In practice, automatically linking SNIDs from different social networks to the same real-world identity is a difficult problem, unless such a link is explicitly available in social network profiles. Note that SNID owners may explicitly reveal their offline identity to the public or to their friends, especially in closed social networks such as Facebook or WeChat, where friends are often connected offline as well. This fact also supports our original claim that the primary goal should be protecting sensitive content instead of protecting identity. In this paper, we consider SNIDs from different social networks as independent – Alice from Twitter and Alice from Facebook are considered unrelated. The rationale is that the privacy protection mechanism from each OSN platform only handles SNIDs, posts, and information flow within its platform. That is, a cross-platform privacy protection mechanism does not exist yet, and discussions of such a mechanism is outside the scope of this paper.

Definition 2.2 (Social Circle). A **social circle (SC)** is simply a set of SNIDs that is owned by a user and used together for some purpose–i.e., SC = $\{\text{SNID}_1, \text{SNID}_2, ..., \text{SNID}_n\}$.

A social circle represents the fact that every social network post is intended for a targeted group of users, where users in the group has inherent social ties or similarities. Posts on similar topics are often intended for the same group. For example, Professor Alice may share research news with *colleagues* ($SC_{Alice,C}$) and *students* ($SC_{Alice,S}$), but only shares her baby's photos with *personal friends*. Note that three different (potentially overlapping) circles are implied in this example. This is similar to the Role-Based Access Control (RBAC) model, where users are assigned to roles, and access rights are associated with roles. In the rest of the paper, we use *Alice* to denote the owner of multiple social circles.

As discussed above, we assume that SNIDs in the same social circle all belong to the same social network. The implication is that we do not concern partially overlapping social circles from different platforms. For example Alice may use her Google+ SNID to interact with dancing buddies (e.g., SC_1={(facebook, xy12), (facebook, yul2)}) and a LinkedIn SNID to interact with professional colleagues (e.g., SC_2={(linkedin, jkim), (linkedin,xza2), (linkedin,xy12)}). In this scenario, they are considered two unrelated SNIDs and three unrelated social circles.

Definition 2.3 (Information Belongings). A user's **information belongings** (I), defined as his or her personal attributes (e.g., birthdate, SSN), content created (e.g., writings, photos), or traces of online social activity (e.g., joining a club, adding a friend).

In our model, each SNID's information belongings are one of the following types: (a) attributes, which often contain sensitive information (e.g., date of birth, SSN); (b) information created or recorded by the person (e.g., writings, photos, videos); (c) public content forwarded or re-posted by the SNID (i.e. Alice posts a news article to her wall); (d) log of social activities (e.g. Alice joins a club, adds a friend); and (e) information generated by a user in response to another information belonging (e.g., Bob replies to Alice's post).

First, all original attributes and messages (type-a and b) from Alice are considered her privacy. Different information content may pose different levels of privacy concern (e.g., a blog post about a national park vs. a blog post about family members). However, as none of the existing social networks provides the capability of autonomous content analysis and content-based privacy protection, all information belongings in each category are subject to the same privacy protection mechanism.

Meanwhile, a type-c information belonging (third-party message) is usually considered as *not* private, however, any comment added by the social circle owner is considered private. Moreover, the fact that the social circle owner forwarded the message is considered a type-d information belonging (a social activity), which is also private. For example, forwarding and commenting on a news report about the presidential election may reveal Alice's political opinions, even though the message itself is non-private.

Last, the type-e information belongings – small pieces of information attached to another (seed) information belonging – are the trickiest. When Bob "likes" or replies to Alice's post, it is both conceptually and practically unclear about the controller of the type-e information belonging. Although the content is generated by Bob (i.e. Bob is the owner), it is impractical for Bob to explicitly specify a social circle for each reply. In practice, different social networks handle them differently, while some implementations introduce potential privacy issues (to be elaborated in Section 3).

Definition 2.4 (Protection Boundary). The **protection boundary** of an element within a user's information belongings is defined as the union of the social circles within which that element has been shared.

The core component of the social circle model is the protection boundary. We assume that when a user posts one or more belongings to a social circle, the individual who owns the social circle and the belongings are one and the same. That says, only Alice

could post an information belonging to her own social circle. Another SNID in Alice's circle, Bob, could "reply to" (or "like") this information belonging, however, when Bob attempts to "forward" the information belonging to his own circle (i.e., outside of Alice's original circle), there may be a potential privacy leakage.

When Alice distributes an information belonging to one of her social circles, she essentially creates a mapping between the information belonging I_i and all the SNIDs in the social circle. This is a many-to-many mapping, which should only be determined by Alice. In practice, the SNIDs who see the information belongings should exactly match Alice's information sharing intention. That is, the *perceived* protection boundary (SC_p), the *specified* protection boundary (SC_s), and the *enforced* protection boundary (SC_e) should all be identical.

2.3 Properties of the Social Circle Model

With the definition of the key elements of the social circle model, now we elaborate the desired properties for a social network privacy protection mechanism. In particular, the protection boundaries that determine who can access the information belongings should be controlled by the user and enforced in a non-leaky manner. At the same time, what the user perceives as his or her protection boundaries should be consistent with what he/she specifies to the social network provider.

Property 1 (Control)**.** The protection boundary of an information belonging should be fully controllable by its owner.

This property states that the social circle owner controls the enforced circle (SC_e). This is the most fundamental property of the social circle model. It implies that nobody, including the social network platform, should violate the information sharing intentions of the owner of the information belonging. Example 1.1 in Section 1 presents a typical violation of this property, in which the user lost control of her protection boundary. When a member in any social circle attempts to move information from its origin to his/her circles, it is the online version of "social gossips", which results a violation of *Property Control*.

The *Control* property provides a theoretical guidance on how "forward" functions should be implemented in social networks. When Alice posts a message to her designated social circle (SC_{Ai}), which includes Bob, Bob should: (1) be disallowed to forward the message; (2) be allowed to forward the message to its original (Alice's) social circle SC_{Ai}; or (3) be allowed to forward the message to a smaller circle, i.e. a subset of SC_{Ai}, to be defined by Bob. Option (2) is the design choice of some social network platforms. Although it does not violate Alice's protection boundary, it will disclose the fact that "Bob is Alice's friend and he forwarded the message" to people who are not Bob's friends (i.e., SC_{Ai} SC_B). If Bob is unaware of this fact and makes a different assumption (that the message is only visible to his friends), it may violate Bob's information sharing intention. Option (3) is usually implemented in this way: Bob is allowed to select a protection boundary (circle SC_B) for the forwarded message, the new boundary to be enforced would be the intersection of two circles $SC_{Ai} \cap SC_B$. This appears to be the best option, which ensures that Alice's information belonging does not escape from its original circle SC_{Ai}, and also gives Bob enough control to his type-d information belonging.

Meanwhile, as we have discussed, type-c information belongings (third party messages such as news articles) are generally not considered to be private. But the corresponding type-d information belongings are private. That says, when Alice forwards a news article to a designated circle (SC_{Ai}), Bob should: (1) be able to forward the original article itself to his designated circle, which is unrelated to Alice or SC_{Ai}; (2) be able to forward the article, with Alice tagged to it, in the same way as Alice's other private information belongings, i.e., same as options (2) and (3) as described above.

Last, for simplicity and usability concerns, there is no explicit control of type-e information belongings, i.e., when Bob replies to Alice's post, he cannot control who sees the reply. One option is to apply the same protection boundary as the seed – whoever sees the seed message sees all the replies. This may violate Bob's information sharing intention, when he does not want non-friends to see his reply. Another model is to implement a *default circle* for every user, so that the reply is visible to the intersection of original circle for this message and Bob's default circle: $SC_{Ai} \cap SC_{B,F}$

Property 2 (Consistency)**.** The user-perceived protection boundary should be consistent with the protection boundary enforced by a social network site

This property reiterates the concepts of the *perceived boundary*, the *specified boundary*, and the *enforced boundary*. To correctly enforce Alice's information sharing intention, we expect $SC_p = SC_s = SC_e$. In particular, when there is inconsistency between perceived and specified boundaries ($SC_p \neq SC_s$), it implies a potential usability issue with the privacy modeling of the social network platform. When there is inconsistency between specified and enforced boundaries ($SC_s \neq SC_e$), it indicates an implementation error, which often causes leaky boundaries.

Property 3 (Usability)**.** The designs of the system should facilitate relatively easy specification and utilization of social circles that are consistent with users' perceptions. The designs of the system should not obscure the scope and extent of socialization and information sharing.

The usability issues occur throughout the design and implementation of social network privacy protection mechanisms. In the social circle model, the usability concerns are: (1) it should be easy to define social circles; (2) the defined social circles should be consistent with user perceptions; (3) it should be easy to select a circle in posting messages. The core of the proposed *Usability* property is to ensure $SC_p = SC_s$ through a user-friendly mechanism.

A major drawback in the adoption of the social circle model is the usability problem–it is tedious and labor-intensive to assign hundreds of existing friends into circles or lists. To tackle this problem, it was proposed that the owner may specify attribute values or credentials so that qualified SNIDs are automatically admitted to the circle. However, such delegation may suffer from discrepancies between user perceptions and specifications. Therefore, undesired SNIDs may be introduced to a circle due to faultily or incompletely specified credentials. There are also proposals to automatically identify social circles based on friendship connections, content and socialization activities (e.g., [60]). Such proposals are not yet adopted in commercial OSN products. Meanwhile, they do not guarantee 100% accuracy – human adjustments are still needed.

Property 4 (Clean Deletion)**.** When the owner of a private information belonging (I) attempts to delete the information belonging from the social circle, I should be completely erased from all users in the social circle.

Information belongings posted to social circles may need to be deleted/recalled, such as the "regretted messages" [54]. The *Clean Deletion* property implies the full control of the social circle owner over the deletion of private information belongings. In addition to the *Control* property, the owner should also have control to remove an information belonging from the social circle. The deletion should be clean that no SNIDs in the circle should see any "phantom post", and no SNIDs in the circle could be able to infer anything about the deleted post.

Different types of information belongings may be deleted differently. In particular, when a type-b information belonging is deleted, all of its instances (e.g. forwarded or re-posted instances) should be completely erased. Meanwhile, its related replies, "likes", and "notifications" (e.g., Bob receives a notice when Alice posts a new message) should be erased too. Meanwhile, there is the concept of "forward privacy" in messages deletions: when Alice posts a message and deletes it before Bob logs in, Bob should be completely unaware of the existence of the deleted message. On the other hand, deleting a type-c information belonging only removes the forwarded copy of the message, but should not affect the original message. Last, if an SNID is closed, all of its information belongings should become invisible.

Property 5 (Non-leaky)**.** The protection boundary of an information belonging should not be leaky.

As a complement to the *Usability* property, the Non-leaky property focuses on the enforcement of the protection boundaries, i.e., to ensure $SC_s = SC_e$. Many social network sites enforce privacy protection as "messages are only accessible within the owner's circle." However, violations of properties 1–4 all result in leaky boundaries. Examples include mal-functioning applications (e.g., Example 1.1 1 in Section 1). In fact, non-leaky models and the non-leaky enforcement are two very different concepts. When pragmatic tradeoffs between usability and privacy have to be done, online social networks may often enforce a non-leaky model in a leaky way. In general, most of the leaky-boundary issues are caused by message forwarding and non-clean deletion. In the next section, we will examine the most popular social network platform on the Internet, and discuss the privacy leakages we discovered from them.

3 ANALYSIS OF SOCIAL NETWORK PRIVACY MECHANISMS

In this section, we use the social circle model to examine the privacy protection functions for information dissemination in six popular social networks. We do not include public social networks, such as Twitter, that do not provide mechanisms to restrict access to user-generated content. For each social network, we mainly focus on the following: (1) The definition (configuration) of social circles, especially the usability issues when adding users into circles; (2) The control of information dissemination into social circles, especially whether the specified and enforced circles are consistent; (3) The security of the protection boundaries, especially, whether the boundaries are leaky when the information belongings are being forwarded; (4) The handling of Type-e information belongings (e.g., likes and replies); and (5) The clean deletion of information belongings. Unless specified otherwise, all the experiments discussed in this section were conducted in December 2017 and January 2018.

3.1 Facebook

Facebook is reported to be the largest online social network platform on the Internet, with 2 billion monthly active users. Facebook started as an internal social networking platform for Harvard College students, and later expanded to more universities and eventually to the public. As the business interest of Facebook is to facilitate socialization and sharing, their privacy policy used to be quite loose, such as: "*The default privacy setting for certain types of information you post on Facebook is set to 'everyone.'*" With years of development, the privacy protection mechanisms in Facebook have evolved significantly.

Defining Social Circles: Initially, Facebook's privacy settings were mostly based on the concept of "networks", which includes schools, geography, etc. For example, when a user registered with an @cmu.edu email, she became part of the "CMU network". Although many users perceive their default protection boundary to be "friends", however, by default, profile attributes and activities were open to their networks (e.g., entire University network) [19]. In this case, everyone has several default circles, but he/she has no control over the membership of such circles (violation of Property 1. Control).

In its currently privacy protection mechanisms, Facebook allows users to organize friends into *custom lists*, which are equivalent to social circles. Lists could be created by adding users one-by-one. Meanwhile, since Facebook explicitly collects user attributes such as location, education, work, etc, it also creates *smart lists* for users (e.g., all friends from CMU are placed into one smart list). Other than the smart lists, users cannot add friends in batch operations.

Control: When a user posts a message, she could choose a protection boundary for the message, including public, friends, friends of friends, or a custom list. This allows a user to define any arbitrary protection boundary for each message. Note that, when a friend is explicitly *tagged* in a message, he/she is automatically added to the enforced circle and cannot be removed from it. However, this function could be confusing that: (1) when a custom list (SC_c) is selected and a friend (f_i) is tagged, the actual protection boundary is the *custom list + anyone tagged*, i.e., $SC_c \cup \{f_i\}$; however, (2) when the default list of *friends* (SC_F) is selected and a friend (f_i) is tagged, the new protection boundary is *friends + anyone tagged + friends of anyone tagged*, i.e. $SC_F \cup f_i \cup SC_{f_i,F}$, unless the user explicitly goes into the custom settings to un-check "friends of anyone tagged". Note that f_i is the owner of circle $SC_{f_i,F}$. Therefore, the user is posting information to someone else's circle, which is out of the control of the user (violation of Property 1. Control). This practice may not seem intuitive/appropriate to all users.

Forwarding: In the current version of Facebook, an information belonging that has restricted access, i.e., NOT available to public, cannot be re-shared/forwarded any more, except for type-c information belongings. Type-c information belongings could always

be re-shared, but the previous sharer's information is excluded in re-sharing. That is, when Alice shares an ESPN news article to her friends, Bob could re-share it, but he is only re-sharing the original article from ESPN, without any indication that it was from Alice. Meanwhile, when Alice posts a Type-b information belonging to *public* and Bob re-shares it, Bob could specify the protection boundary of his re-share. Bob's friend Cathy could re-share from Bob's wall, but she is actually re-sharing Alice's seed information belonging. The fact that Bob re-shared the message (the type-d information belonging) could not be further re-shared.

Replies: In Facebook, all *replies* and *likes* (Type-e information belongings) inherent the protection boundary of the seed content. Therefore, the actual owner of the information belongings has no control over the enforced protection boundary of Type-e information belongings (violation of Property 1. Control). For example, when Bob replies to Alice's message, the enforced protection boundary of the reply ($SC_{e,Bob}(R)$) is the same as the protection boundary of Alice's original message ($SC_{e,Alice}(M)$), which is defined by Alice. It would be interesting to examine if/how Bob's perceived information boundary ($SC_{p,Bob}(R)$) would be different from the enforced boundary $SC_{e,Bob}(R)$. In practice, Bob's reply could be viewed by total strangers of Bob, which could be considered a potential privacy leakage. In the literature, [37] also mentioned that users might be unaware of the possibility of privacy leakage in replies in the context of Twitter.

Clean deletion: Facebook supports clean deletion. In particular, when Alice posts a message that tags Bob, and deletes the message before Bob logs in. Bob will not receive any notification about the message or the tagging. Meanwhile, when a seed information belonging is deleted, all re-shares are also erased.

The Main Takeaway: Facebook, as one of the largest online social networks, has implemented a privacy model that supports social circles. It provides rich functions in defining circles and control protection boundaries. Sometimes there could be inconsistency between perceived and enforced protection boundaries, mainly due to the complications with tagging and the default circle of Friends. The enforced boundaries are non-leaky. In summary, we feel that Facebook's implementation of the social circle model is mostly correct, as long as the users use it correctly.

3.2 Google Plus (Google+)

Google plus, launched in 2011, is an Internet based social network that is operated by Google with 375 million active users as of August 2017. Circles is one of the core functions in Google+.

Define Circles: The default circles in Google+ are friends, family, acquaintances, and following. Users could create new circles and add any arbitrary set of users to any circle. Note that adding a user into a circle, even into the Friends circle, does not require mutual following relationship. Each user needs to be added manually, while there is no mechanism to add a bulk of followers into one circle.

On the other hand, Google+ also has the concept of "Communities" and "Collections". Each community is like a discussion room for people with similar interests. The owner of the community has full control of the membership of the community – the community may be open but the owner could remove unwanted members from the community. Collections are used as a container for posts, where all the posts in the same collection inherit the same protection boundary as the collection.

Control: When a user posts a message, she could choose one of the following as the destination: (1) a community, (2) a collection, or (3) a set of circles and users (followings and followers). In option (3), the enforced protection boundary is the union of all selected circles and users: $SC_e = SC_s = (\bigcup_i SC_i) \cup (\bigcup_j U_j)$. It is not possible to specify other operations such as $(SC_i \cap SC_j)$ or $(SC_F \setminus \{Alice\})$. In options (2) and (3), the user has full control of the protection boundaries. On the other hand, all Type-e information belongings inherit the protection boundary from the seed.

Although communities are not intended for access control or privacy protection, a user could choose a community as the destination when she posts a message (in the same way she chooses circles as destinations). However, the user could post to any community that she is a member of. Hence, she actually does not have any control on who sees the post now and in the future. The fact that communities and circles could be selected in the same way as the destination of a post could be confusing to the users, and may cause potential privacy issues (Potential violation of Property 2. Consistency).

Forwarding: Unless the owner explicitly disables the sharing option for a post, any user is allowed to forward both Type-b and Type-c information belongings that she has access to. For an information belonging that is not posted to public ($SC_e(M) \neq U$), a warning will be displayed when someone attempts to re-share it, but the warning could be ignored. Meanwhile, when $SC_e(M) \neq U$, it cannot be forwarded to the public ($SC_e(M') \neq U$). However, the protection boundary for M' could be any arbitrary set of circles or users defined by the re-sharer, and it could even exclude the original owner. As a result, the original users' information is easily leaked beyond the originally specified protection boundary.

Google+ does not distinguish Type-b and Type-c information belongings in re-sharing. That is, when Alice shares a public news article to a non-public circle, people in the circle will have two options in re-sharing this article: (1) Directly re-share from Alice: this is the same as re-sharing any Type-b information from Alice: the warning will be displayed, and the article can only be re-shared to non-public circles. (2) Click on the original article and re-share from there, so that it could be re-shared to public. For both options, the article displayed on the re-sharer's wall will be the same – although he re-shared from Alice in Option (1), there would be no indication of Alice in the re-shared post.

Clean Deletion: Google plus doesn't support Clean deletion. When the seed information is deleted, the re-shares are not deleted. For instance, when Bob re-shares a post from Alice and later Alice deletes the post, Bob's re-share will still exist, showing that it was from Alice. Interestingly, Bob's re-share is not allowed to be further re-shared if the seed information is deleted. Last, if Bob deletes a comment and undo the '+1' action (similar to 'like' in Facebook) made to Alice's post, the notification of the comment will disappear, but the notification of the '+1' will be kept.

The Main Takeaway: Google+ implements a social circle model, in which users have full control in defining circles and posting to circles. The protection boundary gets leaky when the posts are forwarded by followers. Although the leakage could be prevented through disabling forwarding, usability appears to be an issue. Last,

the Community mechanism is not intended to be used the same way as social circles. However, it gives the users a feeling that it could be used to control information dissemination, but the boundary is out-of-control and leaky.

3.3 VKontakte

VKontakte (VK) is an online social networking service that is very popular in Russian. As of January 2018, VK ranked 17th in Alexafis global top 500 sites, and is 5th among social networking sites. VKontakte is more akin to Facebook.

Defining Social Circles: VKontakte allows users to organize friends into default lists (Best Friends, Co-Workers, Family, University Friends and School Friends) or custom lists. VKontakte also has another function similar to social circles named groups, which can be open (anyone can join), closed (Can join by sending a request or by receiving an invitation) or private (can join on invitation). However, similar to Facebook, this function is not designed for privacy protection.

Control: In VKontakte, the social circles could be used in general privacy settings to define an information boundary for each category of information, such as: "Who can view photos of me", "Who can view the Saved photos album". However, the general privacy settings does not include Type-b information belongings, whose protection boundary needs to be configured for each information belonging. When a user attempts to post a status update, she can only choose between Public ($SC = U$) or Friends (SC_F). When she creates an album as a container for photos, she can select fine-grained protection boundaries such as one or more social circles ($SC = \bigcup SC_i$). In VK, privacy settings are very complicate, for instance, hiding your profile does not hide your birthday – you can only configure your birthday to be shown to public or completely hidden in profile editing. Last, all Type-e information belongings inherit the protection boundary of seed message.

Forwarding: In VK, a user can re-share a public Type-b or Type-c information belonging to: her wall, a community she is in, or via a private message to any friend. For a private wall post that is accessible to her (i.e., she is within the protection boundary), she can share it as a private message to any friend, even the ones who are outside of the protection boundary of the post. Meanwhile, for a private photo, she can share it as a private message or directly on her wall. In the latter case, the photo becomes accessible to anyone who sees her wall, which could be public. In either case, there is no warning regarding the privacy of the original owner of the photo. This shows that the enforced protection boundaries could be easily broken, and the social circle is leaky.

Clean Deletion: In VK, when a wall post or photo has been re-shared and then deleted, all the re-shared copies will continue to exist, and they could be further re-shared. That says, when an information belonging is re-shared, a copy is made and its control belongs to the re-sharer.

The Main Takeaway: VK does not do a good job in protecting user privacy, especially in message forwarding. VK used similar designs with Facebook. The privacy vulnerability during forwarding existed on Facebook in 2014 and earlier. Facebook fixed this vulnerability and enforced tighter privacy restrictions, unfortunately, VK did not follow. Meanwhile, clean deletion is not supported in VK, which means once a post/photo has been forwarded, the original owner completely loses control to the information belonging.

3.4 Sina Weibo

Sina Weibo is a social media platform developed by Sina, with over 361 million active users in the second quarter of 2017. It is considered as a micro-blogging service, which is similar to Twitter. The design is intended to be open to encourage sharing and socialization. For instance, all followings and followers are open to public. However, it added limited privacy protection functions for microblogs.

Defining Social Circles: Sina Weibo has three different mechanisms that manages related users (followers or followings) in groups. Not all of them are designed for privacy protection, or can be used to control information dissemination. (1) Friend circle: "Friends" are defined as followers who are also followed by the owner (i.e. mutual followers). All friends are automatically added into this circle, while the owner could remove friends from or add them back to this circle. She cannot add non-friends into this circle. Each user could have only one *friend circle*, and she has full control over the membership of this circle. This circle could be used for control the boundary of information dissemination. (2) Weibo Groups: They are initially designed as chat groups, but it is possible to post microblogs that are only visible to a group. Any user could create chat groups by adding followers into the chat. Owners have full control over group membership, but she can configure the group to be open or allow members to invite others. Members could leave the group at any time. Members, other than the owner, do not have full control over the membership of the chat groups. (3) Groups of followings: user can group the accounts that she is following into groups, so that she could view mricroblogs from a specified group of followed accounts. This "group" function is designed for managing information consumption, not for information dissemination – users cannot post to such groups.

Control: When a user posts a microblog message, she could choose a protection boundary as: public, friend circle (SC_F), or a chat group. The friend circle is a true social circle that (1) the owner has full control over the membership of the circle, and (2) the enforced protection boundary is exactly the members in this circle $SC_s = SC_e = SC_F$.

Meanwhile, a message could be posted to any Weibo group (chat group) that the user participates. Although this provides a means of information boundary, the mechanism is inconsistent with the social circle model or any other privacy model for social network information dissemination. In particular, a user may post to a group that is not owned by her, and she has no control over the membership of the group. Therefore, the user cannot specify any arbitrary SC_s through Weibo Group. Moreover, since the groups are dynamic and the user does not have full control over group membership, the perceived boundary SC_p could be dramatically different from the actually enforced boundary SC_e. This is a violation of Property Control.

Forwarding: In Weibo, messages maybe forwarded by others, while the forwarding history is maintained. That is, when Bob forwards a message from Alice, his friend Charlie has two options in forwarding this message: (1) forwarding from Bob's wall, so that

the forwarded message will show a "forward chain" like "forwarded from Bob who forwarded from Alice", this chain could be very long in practice; (2) directly forwarding from Alice, so that the chain only shows "forwarded from Alice". A Type-b or Type-C information belonging that is *not* available to public (e.g. a message to SC_F) cannot be forwarded. Meanwhile, when an original information belonging is public, but forwarded by someone to her SC_F, this forwarded message cannot be re-forwarded. However, anyone who has access to any ancestor node in the forwarding chain could still forward from there.

Replies: In Sina Weibo, there are two different notions of (1) who can reply and (2) who can see the replies. The SNID could configure who can reply to a post, such as everyone, followers, etc. Meanwhile, the protection boundary of Type-e information belongings is the intersection of the protection boundary of the seed message and the circle which is chosen by users to determine who could reply.

Clean Deletion: In Sina Weibo, when a type-b information belonging is deleted, all its replies and notifications of replies are erased. However, if the message has been forwarded, then the forwarder's wall will show a message like "this post has been deleted by the original user", but it does not reveal the SNID of the user. However, when Alice forwards a post, Bob re-forwards it, and Alice later deletes her forwarded post, Bob's message will remain on his wall with the complete forwarding chain which shows that it is re-forwarded from Alice. That is, a user can (almost) cleanly delete an original post, but cannot cleanly delete a forwarded post.

The Main Takeaway: Sina Weibo provides one mechanism that implements a limited social circle model. It supports only one circle for each user, who has full control over this circle. The circle is non-leaky, except potential issues with the inherited protection boundaries of the type-e information belongings. On the other hand, although the Chat Group mechanism could be utilized for access control of microblog posts, it was not intended for this function, and it creates issues with Consistency, Usability, and Control.

3.5 QZone

Tencent Qzone is a public social network platform for Tencent's instant messaging software named QQ. It allows QQ users to publish diverse types of content, such as blogs (journals), microblogs ("Shuoshuo"), photo, music, etc. It already had over 606 million monthly active users in the second quarter of 2017.

Defining Social Circles: Users could define custom lists in QZone, which are expected to work as social circles. Users could extract any subset of friends from QQ into custom lists.

Control: First, a user could define a protection boundary for the entire QZone $SC(Z)$, which could be public (must be QQ user to visit), friends, or a subset of friends. Only users allowed by this protection boundary could access the QZone. In the QZone, different privacy protection mechanisms are developed for different media type: (1) The default protection boundary for blogs and microblogs is public. The owner could specify a new protection boundary $SC_s(M)$ by including or excluding only a subset of friends. The actually enforced protection boundary will be $SC_e(M) = SC(Z) \cap SC_s(M)$. (2) Users could set protection boundaries for albums, so that all photos in the album will inherit the same protection boundary $SC_e(Albumn) = SC(Z) \cap SC_s(Album)$.

Last, just like Facebook, all Type-e information belongings such as replies and likes inherit the protection boundary from the seed content.

Forwarding: There are two types of forwarding activities in QZone: "Re-share" and "Reprint". Re-sharing is like creating a link to the original information belonging (except that microblogs are copied). The protection boundary of re-shared information belongings inherits the protection boundary of the re-sharer's QZone, which cannot be re-configured. Meanwhile, reprinting is making a copy of the original message (for blogs and photos), where a new protection boundary could be set by the reprinter.

When a non-public protection boundary is specified for a microblog, then it cannot be re-shared. However, in the case that the specified protection boundary for the message is public ($SC_s(M) = U$) but the QZone is not open to public (i.e. $SC_e(M) = SC(Z) \neq U$), the message could be re-shared. The re-shared microblog breaks the originally enforced protection boundary – whoever could view the re-sharer's QZone sees the re-shared message: $SC_e(M') = SC_{Bob}(Z) \neq SC_e(M)$.

A blog message could always be re-shared, regardless of the originally specified protection boundary. The new protection boundary will inherit the boundary of the re-sharer's QZone ($SC_e(M') = SC_{Bob}(Z)$). The title and the first few lines of the blog post (and thumbnails of images) will be shown in the re-sharer's QZone, with a link to the original blog – a user in $SC_e(M') - SC_e(M)$ could see all these information, but will get an error clicking on the link. A blog message could also be reprinted, so that a new copy of the blog will be created on the reprinter's wall with a new protection boundary specified during reprint. The reprinted blog becomes completely out of the control of its original owner.

Clean Deletion: In Qzone, deleting a blog will not erase its re-shares – the title, abstract and thumbnail will stay on the re-sharer's shared items, but the link will point to an error message. Reprints are not affected when the original blog is deleted – they are not controlled by the original creator. On the other hand, deleting a microblog will erase all its re-shares. However, when a photo without text description is attached to a microblog, the text content of the microblog will be copied to the description. Deleting the microblog (not deleting the photo from the album) will not affect the description of the photo, and it will continue to be visible with the photo.

The Main Takeaway: QZone implements the social circle model, where any arbitrary circle could be specified when the user posts an information belonging. However, the circles become leaky when information belongings are re-shared or reprinted. Meanwhile, when an information belonging is deleted, not all its occurrences are erased. These vulnerabilities could post serious privacy threats to the owners of the information belongings.

3.6 WeChat

WeChat is an instant messenger and social networking software developed by Tencent. Its monthly active users reached 963 million in the second quarter of 2017. In this paper, we focus on the social networking component embedded in WeChat: the *Moments*, in which users may share: (1) a text message with 0 to 9 photos; or (2) a third party resource (link) such as a news article.

Defining Social Circles: Social networks are defined using tags – SNIDs (friends) carrying the same tag are considered in the same circle. Circles may have overlapping SNIDs. Since WeChat integrates instant messenger with social networking, users could add all the friends from group chats to the same circle.
Control: When a Type-b or c information belonging is posted to the "moment", the default protection boundary is *all friends*, i.e., $SC_{Alice} = SC_{Alice,F}$. The user could set a customized protection boundary by *including* only a subset of circles or *excluding* a subset of circles. That is, the specified boundary could be: (1) $SC_s = SC_{Alice,F}$; (2) $SC_s = \bigcup_i SC_{Alice,i}$; or (3) $SC_s = SC_{Alice,F} \setminus (\bigcup_i SC_{Alice,i})$, where $\bigcup_i$ denotes the union of the circles SC_i. No other set operations are supported. In particular, users cannot specify $SC_s = SC_i \cap SC_j$ or $SC_s = SC_i \setminus SC_j$, unless they explicitly define a circle as $SC_i \cap SC_j$.
Forwarding: Users cannot forward others' type-b information belongings, i.e., original messages and pictures. Users could forward Type-c information belongings, regardless of the original access control settings of the information belongings. However, when Bob forwards a Type-c information belonging from Alice's moments to his own moments, he is actually forwarding from the original source – Alice is never associated with the re-post, so that her privacy is not violated.
Replies and Likes (type-e information belongings): In WeChat moments, protection boundary of type-e information belongings cannot be explicitly specified. The enforced circle is the intersection of the specified circle of the seed message and the replier's default circle. For example, Alice posts a photo to the colleagues circle: $SC_s(M) = SC_{Alice,C}$. Bob, who is a member of $SC_{Alice,C}$, comments on the photo. Bob's comment is only visible to Alice's colleagues who are also friends of Bob. That is, the enforced protection boundary of Bob's reply is $SC_e(R) = SC_s(M) \cap SC_{Bob} = SC_{Alice,C} \cap SC_{Bob,F}$. This model is more restricted than Facebook's model for Type-e information belongings, where $SCe(R) = SC_s(M)$.
Clean Deletion: In WeChat, deleting an information belonging from the moment may not erase all its traces. In particular, we found that notifications are not cleanly erased when the corresponding information belongings are deleted. This vulnerability may result undesired recovery of deleted posts (Type-b information belongings) on some versions of WeChat.
(1) Deletion of Type-e information belongings. When Bob likes or comments on Alice's message (i.e., Bob posts a Type-e information belonging), Alice receives a notification of the activity. When Bob unlikes or deletes the comment (i.e. deletes the Type-e information belonging), Alice still sees a notification saying that "the comment has been deleted" – the notification is updated upon the deletion of the original Type-e information belonging, but not deleted correspondingly.
(2) Deletion of Type-b information belonging. Deleting the original message will result the deletion of all attached Type-e information belongings, but not the deletion of corresponding notifications. The notifications become dangling that the seed messages no longer exist. Meanwhile, users may be able to access the deleted message (phantom message) through the dangling notifications. Here we demonstrate the vulnerability.

In our experiments, a WeChat user posted a picture to his moment (WeChat V6.5.18 on iOS), as shown in Figure 2 (a). His friends liked the picture, hence, he received notifications about the event. Later, he decided to delete the picture, as shown in Figure 2 (b). The picture and the likes disappeared, however, the notifications of the likes, including a low-resolution preview of the first picture, still existed, as shown in Figure 2 (c). To make things worse, clicking on dangling notification would lead to a phantom copy of the original message, as shown in Figure 2 (d) – the likes were gone, but the text message and the picture(s) were all shown. We name this the *Dangling notifications and Phantom posts* vulnerability of WeChat moments. This bug does not appear on all versions of WeChat. In our experiments, clicking on the dangling notifications in the Android version of WeChat (V6.5.14) will not load the phantom post. However, clicking on the thumbnail picture of the dangling notification will always load a phantom picture (original size).
The Main Takeaway: WeChat Moments implement a social circle model, in which each user has a default circle consists all her friends. Circles could be extracted from group chats, which is a plus from the usability perspective. Users appear to have full control over their circles and circles are implemented non-leaky. Type-b information belongings cannot be forwarded, which is somewhat restrictive from usability perspective, but it improves privacy. For Type-e information belongings, unlike other social networks that inherits the protection boundary from their seed posts, WeChat utilizes the repliers' default circles to further tighten the boundary for each reply/like. This is more desirable compared with all other social networks we examined. Last, we found a vulnerability in clean deletion, which may be caused by caching.

4 IMPLICATIONS AND DISCUSSIONS

4.1 Implications

Privacy Model. In this paper, we have articulated the social circle model. We show that this model has sufficient rigor and expressiveness to satisfy the needs for controlling the boundary of information dissemination in social networks. It is also easily understandable and highly usable for non-expert users. Several popular OSNs have adopted privacy models that resemble the social circle model. Although the concept of social circles appears to be straight forward, the three-way interactions between *social circles*, *different types of information belongings*, and *operations on the information belongings* may be confusing to both developers and users. In practice, the implementation/enforcement of the social circle model could be problematic, as we have demonstrated in Section 3 of the paper.
Enforcement. From our investigation on six popular social networking sites, we can see that none of them developed a perfectly non-leaky privacy protection mechanism. Some social networking sites implemented more things right, while some made serious mistakes. In particular, Forwarding has been a big challenge – some sites had relaxed restrictions on forwarding to encourage socialization, which resulted in leaky boundaries when a post is forwarded outside its original circle ($SC(M') \setminus SC(M) \neq \Phi$); while some sites enforce more restrictions on forwarding, so that privacy leaks are less likely. Meanwhile, Clean Deletion has been another issue, where we have identified privacy leakages in the models and inconsistencies in the implementations.
Usability. A privacy protection mechanism needs to be used and used properly in order to be effective. In our investigation, we

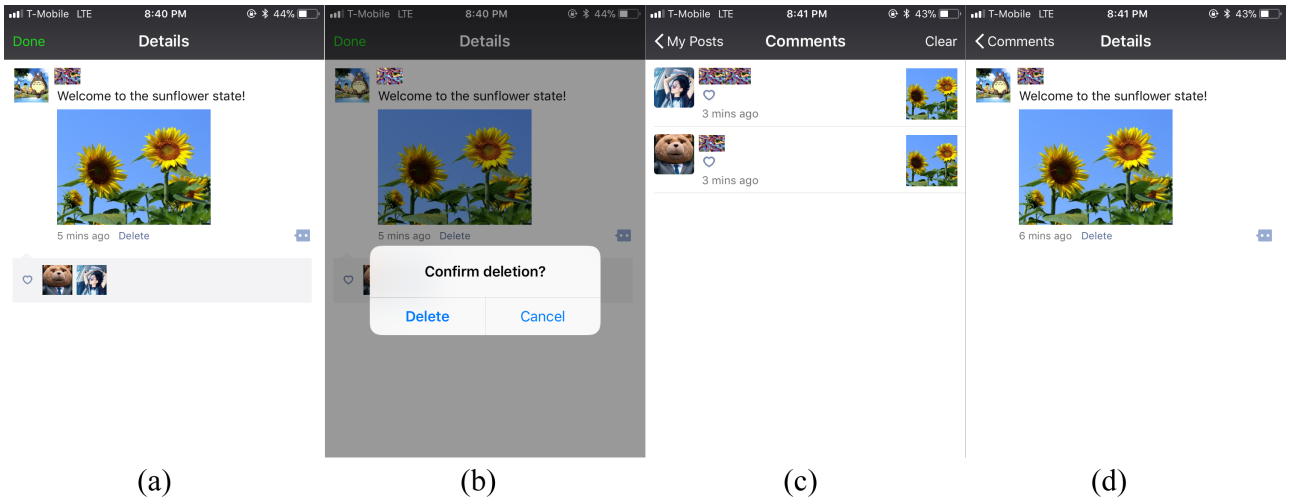

Figure 2: Non-clean deletion in WeChat (this picture is anonymized for double-blind review): (a) User posts a picture, which was liked by two friends; (b) user deletes the message; (c) user could still see notifications that his friends "liked" the picture; (d) user could further click on the notifications to retrieve a phantom copy of the deleted message, but the "likes" were gone.

found that not all privacy protection mechanisms are easily understandable and are convenient to use. For example, in several social networks, the forwarding function needs to be explicitly disabled using a "hidden" checkbox for each post, including posts to private circles. Otherwise the post may be forwarded out of its original circle, without the consent of the original owner of the information belonging. The lack of usability directly leads to privacy threats. Meanwhile, some mechanisms that are not intended for privacy protection (e.g., the Community function in Google+ and the Chat Group function in QZone) also have features similar to social circles that could confuse the users.

4.2 Discussions

Overlapping vs. Non-overlapping Circles. Overlapping circles means two social circles have members in common: $SC_i \cap SC_j \neq \Phi$. Most of the social networks allow overlapping circles since such overlapping exists in real-world relationships, e.g., Bob could be your classmate as well as colleague. However, SNIDs in the intersection of two circles have the privilege to access information disseminated to both circles, which may have the potential to cause privacy threats. However, the models for overlapping circles and non-overlapping circles are convertible: two overlapping circles SC_1 and SC_2 could be translated into three non-overlapping circles: $SC_1 \cap SC_2$, $SC_1 \setminus SC2$, and $SC_2 \setminus SC_1$.

Leaky boundary monitoring. Besides the leaky enforcement of protection boundaries, an active attacker within one social circle could always copy-paste or even make a screen capture of a post, and repost as a new message to a new circle of his/her own. Defending against this attack will require prediction of user behaviors, active monitoring of high-risk nodes (friends who are "gossipy") and content-based detection of similar messages. Such attacks bypass all the privacy protection mechanisms from the SN platform, hence, they are outside the scope of this paper.

Circle Updates. Although social circles are relatively stable – similar to roles in RBAC, they still get updated, especially with newly introduced friends or changes in social status/relationships. When a social circle is reconfigured (by adding/removing members), different social networks take different approaches in managing the protection boundary of existing messages in this circle. For example, Google+ will update the protection boundary to the new circle, while WeChat choose to keep the original protection boundary. That is, when Alice posts a message to a circle and later adds Bob into the circle, Bob sees the message if they were using Google+, but he does not see the message if they were using WeChat. Each of these two design options has its pros and cons, which we cannot provide in-depth discussions here due to space constraints.

Other Perspectives of Social Network Privacy. In the literature, researchers study OSN privacy protection from different angles: (1) protecting user identity in data collection and data publishing (e.g., k-anonymity, differential privacy); (2) access control models and enforcement mechanisms for sharing private information; and (3) preventing users from posting extremely sensitive or regrettable content. In this paper, we follow the second thrust, in which we attempt to identify and evaluate a model that both facilitates information sharing and prevents undesired privacy leakage. We examine the dissemination of potentially private information, and find models and enforcement mechanisms that are able to contain information dissemination as specified by the user.

5 RELATED WORKS

1. Privacy Threats with SN Platforms and Communication. [8] and [64] build social networks from multiple resources while ensuring the privacy of participants. [5] introduces a privacy-preserving social network platform that stores and exchanges encrypted content, and access is enforced through key management. [45] builds a platform that enforces privacy control on third party applications. Meanwhile, users implicitly reveal their identity (e.g.,

IP address) through network communications. Anonymous communications are proposed to hide user identities [11, 17, 43].

2. Privacy Threats within Social Network Sites. *(1) Private information disclosure.* Personal information may be mistakenly disclosed from trusted social networks: publicly-available archives of closed social networks [15], social network stalkers [12], code errors, add-ons and apps [9, 24, 27]. Meanwhile, people publicize private information if they feel "somewhat typical or positively atypical compared to the target group" [23]; 80% of the Facebook users adopt identifiable or semi-identifiable profile photos, and less than 2% made use of the privacy settings [19]. Users' privacy settings violate their sharing intentions [33, 36], and they are unable or unwilling to fix the errors [36]. [28, 44] studies the discrepancies between users' perceived privacy disclosure and the actual exposure allowed by privacy policies. [37] explores three types of private information (e.g., medical conditions) shared in the textual content of tweet messages. Users may also post messages and later regret doing so for various reasons [42, 47, 54]. Impersonation attacks have been proposed [7] to steal private (friends-only) attributes by faking user identities. *(2) Information aggregation attacks.* We introduced information aggregation attacks in [29, 34, 61]: significant amount of privacy is recovered when small pieces of information submitted by users are associated. In particular, people are highly identifiable with very little information [18, 51], which make cross-network aggregations quite feasible. [6] confirms that a significant amount of user profiles from multiple SNSs could be linked by email addresses. *(3) Inference attacks.* Hidden attributes are inferred from friends' attributes with a Bayesian network [21, 22]. Unknown user attributes could be accurately inferred when as few as 20% of the users are known [40]. Friendship links and group membership information can be used to (uniquely) identify users [58] or infer sensitive hidden attributes [65], e.g., membership of a local engineer society discloses user's location and profession [65].

3. Privacy Threats in Published Social Network Data. *(1) Attribute re-identification attacks.* When social network data sets are published for legitimate reasons, user identities are removed. Some well-known techniques include *k-anonymity* [52], *l-diversity* [35] and *t-closeness* [30]. *(2) Structural re-identification attacks.* Graph structure from anonymized social network data could be utilized for re-identification (survey: [67]). Notably, [4] identifies the problem that node identities could be inferred through passive and active attacks. *Topological anonymity* quantifies the level of anonymity using the topology properties [46]. Adversaries with knowledge of user's neighbors could re-identify the user from network graph [66]. *k-degree anonymity* requires each node to have the same degree with at least $k-1$ other nodes [31]. [20] models three types of adversary knowledge that could be used to re-identify vertexes from an anonymized graph. [32] handles social network as a weighted graph, in which edge labels are also considered sensitive.

4. Social Network Privacy Models. With the observation that it is difficult to explicitly define access control for large number of friends, tools have been built to help users manage their privacy settings: *Privacy Wizards* [16] builds a machine learning model to predict and configure users' privacy rules based on limited input, and *PViz* [38] is proposed to help users comprehend their privacy configurations based on the automatically labeled groups. [50] predicts privacy policies for newly uploaded images based on their content similarities with existing images with known policies. Other approaches [59, 62, 63] help users group their contacts, by exploiting the topology relationships among the users' friends. However, none of the above mentioned approaches prevents privacy leakage during normal socialization, and some of them lack theoretical foundations from sociological/psycological perspectives, while others do not have formal constructs.

6 CONCLUSION

With the extreme popularity of online social networks, it is crucial to protect user-generated content that is private or sensitive, without preventing users from normal socialization. In this paper, we articulate the social circle model, which aims to protect the boundary of information dissemination in social networks. We then use this model to examine six popular social networks: Facebook, Google+, VK, Tencent QZone, Weibo, and WeChat. We show that all social network platforms have issues in their implementations of the social circles that may put users' privacy at risk. Some of them pose severe vulnerabilities as their protection boundaries are leaky and sensitive information could flow out of the circle to a significantly larger audience. We also briefly discuss the implications of our findings, and other important issues that are relevant to the social circle model.

REFERENCES

[1] A. Acquisti. 2004. Privacy in Electronic Commerce and the Economics of Immediate Gratification. In *Proceedings of the 5th ACM Electronic Commerce Conference.*

[2] A. Acquisti and J. Grossklags. 2005. Privacy and Rationality in Decision Making. *IEEE Security and Privacy* (2005).

[3] I. Altman. 1975. *The Environment and Social Behavior: Privacy, Personal Space, Territory, and Crowding.* Brooks/Cole Publishing, Monterey, CA.

[4] Lars Backstrom, Cynthia Dwork, and Jon Kleinberg. 2007. Wherefore art thou r3579x?: anonymized social networks, hidden patterns, and structural steganography. In *Proceedings of ACM international conference on World Wide Web.* 181–190. DOI: http://dx.doi.org/10.1145/1242572.1242598

[5] Randy Baden, Adam Bender, Neil Spring, Bobby Bhattacharjee, and Daniel Starin. 2009. Persona: an online social network with user-defined privacy. *SIGCOMM Comput. Commun. Rev.* 39, 4 (2009), 135–146. DOI: http://dx.doi.org/10.1145/1594977.1592585

[6] Marco Balduzzi, Christian Platzer, Thorsten Holz, Engin Kirda, Davide Balzarotti, and Christopher Kruegel. 2010. Abusing Social Networks for Automated User Profiling. In *Recent Advances in Intrusion Detection*, Somesh Jha, Robin Sommer, and Christian Kreibich (Eds.). Lecture Notes in Computer Science, Vol. 6307. Springer Berlin / Heidelberg, 422–441.

[7] Leyla Bilge, Thorsten Strufe, Davide Balzarotti, and Engin Kirda. 2009. All your contacts are belong to us: automated identity theft attacks on social networks. In *Proceedings of the 18th international conference on World wide web (WWW '09).* ACM, New York, NY, USA, 551–560. DOI: http://dx.doi.org/10.1145/1526709.1526784

[8] Gary Blosser and Justin Zhan. 2008. Privacy Preserving Collaborative Social Network. In *International Conference on Information Security and Assurance (ISA).* 543 – 548.

[9] Pete Cashmore. 2009. Privacy is dead, and social media hold smoking gun. CNN. (October 2009).

[10] M.J. Culnan and J.R. Bies. 2003. Consumer Privacy: Balancing Economic and Justice Considerations. *Journal of Social Issues* 59, 2 (2003).

[11] Roger Dingledine, Nick Mathewson, and Paul Syverson. 2004. Tor: the second-generation onion router. In *USENIX Security Symposium.*

[12] Byron Dubow. 2007. Confessions of 'Facebook stalkers'. USA Today. (March 2007).

[13] Cynthia Dwork. 2008. Differential privacy: A survey of results. In *International Conference on Theory and Applications of Models of Computation.* Springer, 1–19.

[14] N.B. Ellison, C. Steinfield, and C. Lampe. 2007. The Benefits of Facebook "Friends:" Social Capital and College Students' Use of Online Social Network Sites. *Journal of Computer-Mediated Communication* 12, 4 (2007), 1143–1168.

[15] Gunther Eysenbach and James E Till. 2001. Ethical issues in qualitative research on internet communities. *BMJ* 323 (2001), 1103–1105.

[16] Lujun Fang and Kristen LeFevre. 2010. Privacy Wizards for Social Networking Sites. In *International World Wide Web conference(WWW)*.

[17] David Goldschlag, Michael Reed, and Paul Syverson. 1999. Onion Routing for Anonymous and Private Internet Connections. *Commun. ACM* 42, 2 (1999), 39–41. DOI: http://dx.doi.org/10.1145/293411.293443

[18] Philippe Golle. 2006. Revisiting the uniqueness of simple demographics in the US population. In *WPES '06: Proceedings of the 5th ACM workshop on Privacy in electronic society*. ACM, New York, NY, USA, 77–80. DOI: http://dx.doi.org/10.1145/1179601.1179615

[19] Ralph Gross, Alessandro Acquisti, and III H. John Heinz. 2005. Information revelation and privacy in online social networks (The Facebook case). In *Proceedings of ACM workshop on Privacy in the electronic society*. 71–80. DOI: http://dx.doi.org/10.1145/1102199.1102214

[20] Michael Hay, Gerome Miklau, David Jensen, Don Towsley, and Philipp Weis. 2008. Resisting structural re-identification in anonymized social networks. *Proc. VLDB Endow.* 1, 1 (2008), 102–114. DOI: http://dx.doi.org/10.1145/1453856.1453873

[21] Jianming He and Wesley W. Chu. 2008. Protecting Private Information in Online Social Networks. In *Intelligence and Security Informatics*. 249–273.

[22] Jianming He, Wesley W. Chu, and Zhenyu Liu. 2006. Inferring Privacy Information from Social Networks. In *IEEE International Conference on Intelligence and Security Informatics*. 154–165.

[23] Bernardo A. Huberman, Eytan Adar, and Leslie R. Fine. 2005. Valuating Privacy. *IEEE Security and Privacy* 3, 5 (2005), 22–25.

[24] M Irvine. 2008. Social network users overlook privacy pitfalls. USA Today. (April 2008).

[25] Haiyan Jia and Heng Xu. 2016. Autonomous and interdependent: Collaborative privacy management on social networking sites. In *ACM CHI*.

[26] Sanjay Kairam, Mike Brzozowski, David Huffaker, and Ed Chi. 2012. Talking in circles: selective sharing in Google+. In *ACM CHI*.

[27] Balachander Krishnamurthy and Craig E. Wills. 2009. On the leakage of personally identifiable information via online social networks. In *WOSN '09: Proceedings of the 2nd ACM workshop on Online social networks*. ACM, New York, NY, USA, 7–12. DOI: http://dx.doi.org/10.1145/1592665.1592668

[28] Yann G. Le Gall, Adam J. Lee, and Apu Kapadia. 2012. PlexC: a policy language for exposure control. In *Proceedings of the 17th ACM symposium on Access Control Models and Technologies (SACMAT '12)*. ACM, 219–228. DOI: http://dx.doi.org/10.1145/2295136.2295174

[29] Fengjun Li, Jake Y. Chen, Xukai Zou, and Peng Liu. 2010. New Privacy Threats in Healthcare Informatics: When Medical Records Join the Web. In *Proceedings of the 9th International Workshop on Data Mining in Bioinformatics (BIOKDD)*.

[30] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. 2007. t-Closeness: Privacy Beyond k-Anonymity and l-Diversity. In *Proceedings of the 23rd International Conference on Data Engineering*. 106–115.

[31] Kun Liu and Evimaria Terzi. 2008. Towards identity anonymization on graphs. In *Proceedings of the 2008 ACM SIGMOD international conference on Management of data*.

[32] Lian Liu, Jie Wang, Jinze Liu, and Jun Zhang. 2008. *Privacy Preserving in Social Networks Against Sensitive Edge Disclosure*. Technical Report CMIDA-HiPSCCS 006-08. University of Kentucky.

[33] Yabing Liu, Krishna P. Gummadi, Balachander Krishnamurthy, and Alan Mislove. 2011. Analyzing facebook privacy settings: user expectations vs. reality. In *Proceedings of the 2011 ACM SIGCOMM conference on Internet measurement conference (IMC '11)*. ACM, New York, NY, USA, 61–70.

[34] Bo Luo and Dongwon Lee. 2009. On Protecting Private Information in Social Networks: A Proposal. In *IEEE ICDE Workshop on Modeling, Managing, and Mining of Evolving Social Networks (M3SN)*.

[35] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. 2007. L-diversity: Privacy beyond k-anonymity. *ACM Trans. Knowl. Discov. Data* 1, 1 (2007), 3. DOI: http://dx.doi.org/10.1145/1217299.1217302

[36] M. Madejski, M. Johnson, and S. M. Bellovin. 2011. *The failure of online social network privacy settings*. Technical Report CUCS-010-11. Columbia University.

[37] Huina Mao, Xin Shuai, and Apu Kapadia. 2011. Loose tweets: an analysis of privacy leaks on twitter. In *Proceedings of the 10th annual ACM workshop on Privacy in the electronic society (WPES '11)*. ACM, New York, NY, USA, 1–12.

[38] Alessandra Mazzia, Kristen LeFevre, and Eytan Adar. 2012. The PViz comprehension tool for social network privacy settings. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*. Article 13, 12 pages.

[39] Julian McAuley and Jure Leskovec. 2012. Discovering social circles in ego networks. *TKDD'13* (2012).

[40] Alan Mislove, Bimal Viswanath, Krishna P. Gummadi, and Peter Druschel. 2010. You are who you know: inferring user profiles in online social networks. In *Proceedings of the third ACM international conference on Web search and data mining (WSDM '10)*. ACM, 251–260. DOI: http://dx.doi.org/10.1145/1718487.1718519

[41] P.A. Norberg, D.R. Horne, and D.A. Horne. 2007. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs* 41, 1 (2007).

[42] Sameer Patil, Greg Norcie, Apu Kapadia, and Adam J Lee. 2012. Reasons, rewards, regrets: Privacy considerations in location sharing as an interactive practice. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*. ACM, 5.

[43] Michael K. Reiter and Aviel D. Rubin. 1998. Crowds: anonymity for Web transactions. *ACM Trans. Inf. Syst. Secur.* 1, 1 (1998), 66–92. DOI: http://dx.doi.org/10.1145/290163.290168

[44] Roman Schlegel, Apu Kapadia, and Adam J. Lee. 2011. Eyeing your exposure: quantifying and controlling information sharing for improved privacy. In *Proceedings of the Seventh Symposium on Usable Privacy and Security (SOUPS '11)*. ACM, Article 14, 14 pages. DOI: http://dx.doi.org/10.1145/2078827.2078846

[45] Kapil Singh, Sumeer Bhola, and Wenke Lee. 2009. xBook: Redesigning Privacy Control in Social Networking Platforms. In *Proceedings of 18th USENIX Security Symposium*.

[46] Lisa Singh and Justin Zhan. 2007. Measuring Topological Anonymity in Social Networks. In *GRC '07: Proceedings of the 2007 IEEE International Conference on Granular Computing*.

[47] Manya Sleeper, Justin Cranshaw, Patrick Gage Kelley, Blase Ur, Alessandro Acquisti, Lorrie Faith Cranor, and Norman Sadeh. 2013. i read my Twitter the next morning and was astonished: a conversational perspective on Twitter regrets. In *Proceedings of the 2013 ACM annual conference on Human factors in computing systems*. ACM, 3277–3286.

[48] Anna Squicciarini, Sushama Karumanchi, Dan Lin, and Nicole DeSisto. 2013. Identifying hidden social circles for advanced privacy configuration. *Computers & Security* (2013).

[49] Anna Squicciarini, Dan Lin, Sushama Karumanchi, and Nicole DeSisto. 2012. Automatic social group organization and privacy management. In *CollaborateCom*.

[50] Anna Cinzia Squicciarini, Smitha Sundareswaran, Dan Lin, and Josh Wede. 2011. A3P: adaptive policy prediction for shared images over popular content sharing sites. In *Proceedings of the 22nd ACM conference on Hypertext and hypermedia*.

[51] Latanya Sweeney. 2000. Uniqueness of Simple Demographics in the U.S. Population. (2000).

[52] Latanya Sweeney. 2002. k-anonymity: a model for protecting privacy. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.* 10, 5 (2002), 557–570. DOI: http://dx.doi.org/10.1142/S0218488502001648

[53] H.T. Tavani. 2007. Philosophical Theories of Privacy: Implications for an Adequate Online Privacy Policy. *Metaphilosophy* 38, 1 (2007).

[54] Yang Wang, Gregory Norcie, Saranga Komanduri, Alessandro Acquisti, Pedro Giovanni Leon, and Lorrie Faith Cranor. 2011. I regretted the minute I pressed share: A qualitative study of regrets on Facebook. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*. ACM, 10.

[55] Jason Watson, Andrew Besmer, and Heather Richter Lipford. 2012. + Your circles: sharing behavior on Google+. In *ACM SOUPS*.

[56] A.F. Westin. 1967. *Privacy and Freedom*. Atheneum, New York.

[57] Pamela Wisniewski, Heng Xu, and Yunan Chen. 2014. Understanding user adaptation strategies for the launching of facebook timeline. In *ACM CHI*.

[58] Gilbert Wondracek, Thorsten Holz, Engin Kirda, and Christopher Kruegel. 2010. A Practical Attack to De-anonymize Social Network Users. In *Security and Privacy (SP), 2010 IEEE Symposium on*. 223 –238. DOI: http://dx.doi.org/10.1109/SP.2010.21

[59] Qian Xiao, Htoo Htet Aung, and Kian-Lee Tan. 2012. Towards ad-hoc circles in social networking sites. In *Proceedings of the 2nd ACM SIGMOD Workshop on Databases and Social Networks (DBSocial '12)*. 19–24.

[60] Yuhao Yang, Chao Lan, Xiaoli Li, Bo Luo, and Jun Huan. 2014. Automatic social circle detection using multi-view clustering. In *ACM CIKM*.

[61] Yuhao Yang, Jonathan Lutes, Fengjun Li, Bo Luo, and Peng Liu. 2012. Stalking Online: on User Privacy in Social Networks. In *ACM Conference on Data and Application Security and Privacy (CODASPY)*.

[62] H. Yildiz and C. Kruegel. 2012. Detecting social cliques for automated privacy control in online social networks. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2012 IEEE International Conference on*. 353 –359.

[63] A.S. Yuksel, M.E. Yuksel, and A.H. Zaim. 2010. An Approach for Protecting Privacy on Social Networks. In *Systems and Networks Communications (ICSNC), 2010 Fifth International Conference on*. 154 –159. DOI: http://dx.doi.org/10.1109/ICSNC.2010.30

[64] Justin Zhan, Gary Blosser, Chris Yang, and Lisa Singh. 2008. Privacy-Preserving Collaborative Social Networks. In *Pacific Asia Workshop on Intelligence and Security Informatics (PAISI)*.

[65] Elena Zheleva and Lise Getoor. 2009. To Join or not to Join: The Illusion of Privacy in Social Networks with Mixed Public and Private User Profiles. In *18th International World Wide Web conference (WWW)*. Earlier version appears as CS-TR-4926.

[66] Bin Zhou and Jian Pei. 2008. Preserving privacy in social networks against neighborhood attacks. In *Proceedings of the 24th International Conference on Data Engineering (ICDE)*.

[67] Bin Zhou, Jian Pei, and WoShun Luk. 2008. A brief survey on anonymization techniques for privacy preserving publishing of social network data. *SIGKDD Explor. Newsl.* 10, 2 (2008), 12–22. DOI: http://dx.doi.org/10.1145/1540276.1540279