Preserving User Privacy from Third-party Applications in Online Social Networks

Yuan Cheng
ycheng@cs.utsa.edu

Jaehong Park
jae.park@utsa.edu

Ravi Sandhu
ravi.sandhu@utsa.edu

Institute for Cyber Security
University of Texas at San Antonio
San Antonio, TX, USA

ABSTRACT

Online social networks (OSNs) facilitate many third-party applications (TPAs) that offer users additional functionality and services. However, they also pose serious user privacy risk as current OSNs provide little control over disclosure of user data to TPAs. Addressing the privacy and security issues related to TPAs (and the underlying social networking platforms) requires solutions beyond a simple all-or-nothing strategy. In this paper, we outline an access control framework that provides users flexible controls over how TPAs can access user data and activities in OSNs while still retaining the functionality of TPAs. The proposed framework specifically allows TPAs to utilize some private data without actually transmitting this data to TPAs. Our approach determines access from TPAs based on user-specified policies in terms of relationships between the user and the application.

Categories and Subject Descriptors


General Terms

Security

Keywords

Privacy, Online Social Networks, Social Applications

1. INTRODUCTION

To offer richer functionality, many online social networks (OSNs) have launched social networking platforms that enable third-party developers to contribute applications to the social network through the use of APIs (application programming interfaces). Without the support of OSN platforms, third-party applications (TPAs) have become highly popular in a very short period of time.

The emergence of TPAs also poses severe privacy risks to users. TPAs need to consume users’ and their friends’ data to provide extra functionality to users. Under the current circumstance, TPAs usually receive privileges equal to the TPA users with respect to social graph traversal, and thereby gain access to an abundance of users’ information regardless of the actual legitimate needs. Moreover, these applications are available via OSNs but are running on external servers outside the OSN’s control. Once they acquire the data, they can use or dispose it in whatever way they want without user or OSN consent. There is no control regarding the usage of user data once it is released to the TPA. The developers of TPAs can aggregate such data and accrue benefit by using or selling the data. The most common approach adopted by existing OSNs is a simple all-or-nothing strategy: a user must agree to allow the application to access some subsets of her profile information before installing and using it. Thus, it is difficult for users to know and control how their information is accessed by those various external parties. The only choice for the user to not provide such data is to not use the application. In particular, TPA providers access the contact list of the TPA users and fetch information about the friends of the users, even though those friends did not install the application themselves or consent to this access.

Most previous research on access control in OSNs concentrates on access between regular users in the system, leaving the issues of TPAs out of scope. However, TPAs do not behave like a regular user. They have the ability to aggregate a huge volume of information from their users and can do anything with the collected data without any controls or consent from users or OSNs. We believe OSNs need effective means to prevent privacy leakage for users, in addition to the terms of services and limitation of their APIs. This is likely to be in both OSNs’ and TPAs’ interests as more users are likely to use privacy preserving OSNs and TPAs.

In this paper, we address the issue of inappropriate exposure of user’s information to TPAs. We present an access control framework that provides flexible and fine-grained controls on how TPAs can access OSN user’s data. For this purpose, we classify TPAs into three categories: running outside of OSN, running inside of OSN, and hybrid where some modules are running inside of OSN while others are running outside. The fundamental idea of our approach is to constrain and consume the private information within the OSN system but only allowing privacy-nonsensitive information to be sent outside the OSN for necessary functionality. Therefore, applications or functions running outside the OSN may only receive privacy-nonsensitive data, while those running inside may consume raw private data under the surveillance of the OSN system but are not allowed to transmit the data outside. We also define a relationship-based access control policy language to the framework that allows users to specify how TPAs can access their data in terms of the relationships among users and applications.

2. BACKGROUND

This section provides an overview of OSN platforms, describes the privacy issues in those platforms, and reviews the previous literature on these issues.
2.1 Social Networking Platforms

Many of the popular OSNs have released web APIs to allow third-party developers and websites to implement their own services, which can utilize and aggregate user information and activities in OSNs. Figure 1 presents a typical architecture of current social networking platforms, where TPAs are designed in accordance with APIs and can access user data through APIs. Typically, these applications are integrated in an OSN site but rely on their own external server for running the application.

The Facebook Platform and Google’s OpenSocial are two of the leading forces in the business of social networking platforms. The Facebook Platform [1] is a proprietary software environment for launching TPAs in Facebook. The core of Facebook Platform is the Graph API, which forms the primary way of retrieving and posting data in Facebook. It allows developers to define objects and actions in the social graph, and to create new instances of objects and actions. Google’s OpenSocial [2] is an open source cross-platform competitor to the Facebook Platform for building social applications. It defines a common API for applications across multiple websites, through which OSNs can grant applications access to the social graph as well as messaging service and update feeds. Its greatest advantages over Facebook Platform is its interoperability within the context of multiple OSNs.

2.2 Privacy Issues with Current Platforms

On current platforms, for an OSN user to install an application, a dialog is displayed to her showing that the application requests an access to a list of her profile data. Different from access control policies for user-to-user interactions, the user cannot express her privacy preferences for data that are accessed by applications. In fact, many OSNs currently only provide an all-or-nothing policy when it comes to application-to-user interactions, without letting OSN users specify which information TPAs can access. Thus, even if the application only needs one piece of profile data, the user has to agree to grant the application full access to her profile data; otherwise, she cannot use the application at all. Recently, some OSNs offer users more customized privacy management regarding TPAs than before, where users are allowed to opt-in or opt-out some categories of profile data. However, the essential problem still remains. First, the coarse-grained privacy control does not support users to specify access control policies for each piece of data users own, nor does it distinguish different types of actions application can exercise on data. Second, users are not aware of what kind of data the application really needs. Applications can still ask for more data than they really need, which simply violates the principle of least privilege. We believe it is more appropriate to mediate access on a per-request basis rather than approve access in advance during the time of installation. Third, some permissions are given by user’s friend who installed the application, without user’s knowledge. This poses a serious exposure to those users who are not willing or interested to use the applications.

2.3 Related Works

There has been considerable work seeking to resolve the privacy issues regarding TPAs. Table 1 summarizes the characteristics of some solutions proposed recently.

All communication between applications and data in OSNs is encapsulated by APIs. However, current APIs are not designed with the purpose of access control in mind, letting TPAs access user data without obstruction. To prevent TPAs from directly interacting with user data through APIs, several proposed proxy designs intercept all requests from TPAs, exert users’ privacy preference on data retrieved from the OSN, and then return the sanitized or dummy data to TPAs [3, 8, 9]. The enforcement of such designs varies from server-side proxy to client-side plug-in, according to the trusted entities the designs rely on.

Singh et al [14] presented xBook framework to prohibit untrusted applications from leaking users’ private information. TPAs are hosted on trusted xBook platform, which confines application execution and mediates information flow between different application components. Inspired by xBook, the system proposed by Viswanath et al [15] enforces sandbox structures on both server and client sides to restrict information flow from users to application developers. Our work shares a common design strategy with these two systems by separating applications into multiple components and restricting the information flow between components.

A number of researchers seek to support user-specified privacy preference over what information can be accessed by TPAs. The collaborative privacy management (CPM) framework [3] allows...
users to define privacy configurations and share them with others. In the approach presented by Shehab et al [13], users can choose to opt-in, opt-out, or generalize the application requested data to reflect their access preferences for the applications. PoX [8] uses a dedicated Facebook application to store access control lists for users, which indicate what application should be allowed to access what pieces of information. These proposals contrast with our work in that they do not offer a complete access control policy model for users to specify their privacy preferences. Besmer et al [4] introduced a new application-to-user policy that restricts application’s access to user information while still allowing desirable functionality, which is most similar to ours. However, in addition to the policy model, we also provide a framework that separates application components for finer-tuned information flow controls.

3. THE FRAMEWORK

In this section, we present an access control framework that enforces better privacy preservation for users against third-party applications. Our goal is to prevent TPAs from learning user’s private information while still maintaining the functionality of the applications.

3.1 Overview

As discussed earlier, existing access control mechanisms give applications too much freedom in accessing user’s information. A survey conducted by Felt et al [9] indicated that 91% of the 150 top Facebook applications have unnecessary access to user’s private data, violating the principle of least privilege. On the other hand, completely limiting the applications from user data or providing fake data to them may harm the functionality of applications or even their own business models.

To achieve a balance between application functionality and user privacy, we first present a taxonomy of user’s information based on the value for applications and users, as shown in Table 2. User information can be classified into four categories. The first row indicates data that is inessential for application but of significance to user privacy. It is definitely unnecessary for TPAs to acquire such data, thus requests for the data will not be permitted in any circumstance. If the data is not necessary for running of application and is not private (Row 2), disclosing it does not breach privacy and the decision is up to user’s own choice. For data essential for application, however, we need a special scheme to ensure the functionality without privacy leakage. Unlike other approaches that do not trust the OSN systems [12], we assume the OSN systems are always trustworthy with respect to user privacy, but TPA server may not be. User’s private information should be kept away from these untrustworthy external servers. Among the data that is essential for application functionality, some are of less privacy concerns and thus accessible by untrustworthy external entities (Row 3), while others are sensitive private information that require proper protection against external entities (Row 4).

The premise of our idea is that running an application may not require users’ private information, and even if it requires it not every function of the application needs the private information. There-
3.3 The Application Components

As mentioned earlier, in this framework, the majority of TPAs are likely to run in hybrid form, consisting of both internal and external components. The internal and external components of an application are hosted on servers with different trustworthiness, and can access various user information with different privacy sensitivity. Hence, the communications between these components needs to be precisely managed to prevent unnecessary information leakage. In the proposed social networking platform, the internal and external components can communicate with each other only through OSN-specified APIs which allows only certain types of pre-defined communications.

Between the internal and external components, we define three types of communications: communications with system calls, non-private data and private data. The communication using system calls is a mechanism used by both internal and external functional modules for sending a message (e.g., a function call) to other components. It is used to trigger and handle specific events, and does not contain information about users. In this framework, we allow two-way system call communications between the components. A non-private data communication refers to a message transmission between application components where the message contains non-sensitive information fetched from users, such as keystrokes, mouse clicks, or puzzle results. This type of communication needs to be used if the messages that need to be shared cannot be transmitted using a system call. A communication with private data, on the other hand, is a message transmission between internal and external components that includes users’ private information.

Note that internal modules typically require to access users’ private data as that is why they reside inside the OSN. As internal components reside in the trusted OSN servers, users are more likely to allow the internal components to access users’ private information. Here, we assume all internal modules are evaluated by the hosting OSN and their functional behaviors (such as what kind of data can be released to external components) are verified. As shown in Figure 2, the internal components can be further divided into four types of modules, based on the criteria shown in Table 3. In Table 3, internal modules are classified using two main criteria. The columns show that internal modules can be created/provided by either OSN provider or TPA developer, while the rows show two types of communications based on what kind of messages can be transmitted.

**Module Type 1 (M1):** OSN systems may offer their own functional modules to fulfill some tasks for applications, such as gadgets and widgets. This type of module is trusted as it is created by the OSN system, and thus most likely to be allowed to access users’ private information. It can communicate with other application components through system calls, but any communication related to user data is disallowed.

**Module Type 2 (M2):** Just like M1, this type of module can receive and send system calls with the components on external servers, but is not allowed to transmit any customized messages to the components outside. The difference between M2 and M1 is that M2 is provided by third-party developers rather than the OSN system. This can be the case when an application needs an internal module that is not available by OSN but can communicate with an external component using only system calls provided by OSN API.

The above two modules can be used to perform tasks in which users’ private information is needed and communication between internal and external components can be done using system calls only. One motivating example is the notification sending scenario, where an external component triggers a request, and then an internal component can take over and fulfill the job and request job status information back to the external component without letting the external component know the information about the actual notification recipients.

**Module Type 3 and 4 (M3, M4):** In addition to system call, these two types of module can send privacy-nonsensitive information to the external parties. For example, such modules can record users’ keystrokes and mouse clicks during the game, and send them back to external server. While a sanitized form of users’ private data could be considered as non-private data and belong in these types, sanitization issues are outside of this paper’s scope.

We have identified four possible types of modules for the internal component. While application developers can choose any combination of them to construct the internal component of their application, it is also possible to see these types in a hierarchy as M1 and M2 could be considered more privacy preserving than M3 and M4 depending on the available system calls. For simplicity, we do not consider any ordering of the modules. The proposed social networking platform can mediate information flow in any overt channel between the components, according to user’s policies. While we recognize there remains a covert channel problem, we deliberately keep this issue outside the scope of this paper which is focused on overt communication.

Theoretically, while there could be modules that do not need to communicate with external components, we believe this type of internal modules are less realistic hence not discussed in this paper. Likewise, one can also consider internal modules that can overtly transmit users’ private data to TPA server. However, this approach does not preserve user privacy, hence is not considered in this paper.

4. THE POLICY MODEL

Most OSNs offer access control policies for regulating user-to-user interactions, which are usually based on relationships between users in the social graph. Many researchers including ourselves also proposed a number of relationship-based access control [11] solutions that can improve the current approach from different aspects [5–7, 10]. At the same time, the prevailing approach of addressing application-to-user interactions in existing social networking platforms is rudimentary and coarse-grained to say the least. For example, in Facebook, users now can select what kind of information they are willing to share or not share with third-party applications in general. But for a specific application, the only thing users can do is to approve all access the application requests during the installation stage, or otherwise deny the request and terminate the installation.

The most unique characteristic that distinguishes OSNs from other systems is that users and data objects in OSNs are interconnected through different types of relationships in the social graph. Installing and using an application can be also viewed as an establishment of a special relationship between user and application. Hence, it is intuitive to apply relationship-based access control to govern application-to-user interactions. Inspired by our previous work about user-to-user interactions [6, 7], we propose a policy model for controlling application’s access in OSNs in terms of relationships.

<table>
<thead>
<tr>
<th>Table 3: Module Types in Internal Components</th>
</tr>
</thead>
<tbody>
<tr>
<td>Communication w/ system calls</td>
</tr>
<tr>
<td>OSN provided</td>
</tr>
<tr>
<td>M1</td>
</tr>
<tr>
<td>Communication w/ non-private data</td>
</tr>
<tr>
<td>M3</td>
</tr>
</tbody>
</table>
4.1 Relationship-based Policies

Social graph depicts the relationships between users in OSNs, and it can be extended to include resources as well. Information sharing and social interactions among users are typically based on such graph. Social networking platforms would also benefit from the social graph by taking advantage of the social relationships to offer a richer experience for users. The platform would allow an application to traverse the social graph to access information owned by the application user’s friends through the application user. Thus, applications are able to reach users that have not installed the application at all.

In most existing OSNs, access control for user-to-user interactions is based on the topology of the social graph, where granting access permission is subject to the existence of a particular relationship or a sequence of relationships between the access requester and the target or the owner of the target, and the access control policies are specified in terms of such relationships. We can apply this paradigm of access control to application-to-user interactions by connecting applications to the social graph and considering application-to-user relationship for relationship-based access control.

Our framework comes with a policy model that allows users to specify how applications can access information owned by themselves and their friends in terms of the relationships between the application and the users. The available relationships for users to choose from include “install” relationship between the application and the users, in addition to all other user-to-user relationships among users in OSNs. Basically, users can use a combination of these relationships to govern application’s access to resources.

As mentioned earlier, applications can access the friends’ information of an application user without consent from the friends, and sometimes, beyond the application user’s expectation as well. Effectively, users have delegated the rights to decide which applications can be trusted to their friends. To resolve the problem, some OSN sites now offer users the ability to select categories of information that can be disclosed to applications that they have not installed in case their friends use those applications. However, this mechanism is still far from user’s expectation for fine-grained access control. To this end, the proposed framework considers policies of both the application users and their friends regarding the application’s access. Although most current APIs only allow applications to access direct friends of their users, our policy language actually supports the extension to multiple hops of relationships for more expressive policies.

4.2 Policy Specification

We model an application’s access request to user information as a tuple 
\[
\langle \text{requester}, \text{action}, \text{target} \rangle,
\]
where \(\text{requester}\) indicates the application that launches the access request, \(\text{action}\) denotes the type of access the requester wants to exercise, whereas \(\text{target}\) represents the target object of the access. A \(\text{target}\) can be either the user executing the application (i.e., application user), the application user’s friend, or a data object owned by one of them. The set of supported actions is determined by the API of the social networking platform. The granularity of target objects is chosen based on the system designer’s decision, possibly varying from data object to data type. Access control policies for application-to-user interactions are composed of four elements:

\[
\langle \text{action}, \text{target}, (\text{start, path rule}), 2^{\text{ModuleType}} \rangle,
\]

where \(\text{action}\) specifies the type of access, and \(\text{target}\) is an optional parameter only for access against resource, denoting the resource to be accessed. It remains blank when the requested access is against a user.

For the third element, \(\text{start}\) is the position where access evaluation begins, which can be either the target user indicated in the request, the owner of the target object (denoted as \(\text{owner}\)), or the application requesting access (denoted as \(\text{requester}\)). A \(\text{path rule}\) is composed of one or more relationship paths, with each representing the required pattern of relationship between the involved parties in order to grant access. For example, \(\text{install}\) represents the relationship between the application and the application user, “\(\text{install-friend}\)” denotes the relationship between the application and the friend of the application user, whereas “\(\emptyset\)” indicates no explicit relationship path but the starting node itself is allowed. Given a pair \((\text{start}, \text{path rule})\), if \(\text{start}\) is the target user or resource owner, the reference monitor checks if the requester (i.e., application) can be reached via relationship path specified in \(\text{path rule}\). If \(\text{start}\) is the requester, the reference monitor then evaluates from the requester to see the relationship path between the requester and the target user or resource owner.

The last parameter \(2^{\text{ModuleType}}\) indicates the set of application module types that is allowed to access, where \(\text{ModuleType}\) is a set composed of \(M1, M2, M3, M4\) and \(\text{external}\). It enables the resource owner to specify different policies for different application modules regarding the same pair of action and target. For example, user Alice may allow some internal modules of an application (see App 3b in Figure 2) she installed to access her birthday, but disapproves the external component of the application to access (see App 3a in Figure 2) the same data.

4.3 Examples

Below we show two examples to elaborate how users can use the policy model to control application’s access in OSNs.

Example 1: App Request Notification. Once a user has installed and used an application, the application may start sending app requests to the user’s friends or suggesting new applications to the user. The notification of such requests is annoying to many people. In general, users usually tend to hide their identities from applications they do not know. In this case, the following policies may help user Alice and her friends keep their privacy with respect to app requests:

- For applications she installed #1: <app request, \(\_\) (target user, install), \(\{M1, M2, M3, M4, \text{external}\}\) >
- For applications she installed #2: <app request, \(\_\) (requester, install-friend), \(\{M1, M2\}\) >
- For applications her friends installed: <app request, \(\_\) (target user, friend-install), \(\{M1, M2\}\) >

The first policy applies to requests such as < AppX, app request, Alice >, saying that applications she installed are allowed to send her app requests about activity updates and suggestions for new applications. The pair \((\text{target user}, \text{install})\) expresses the applications that she installed. The second policy states that only the M1 or M2 modules of the applications she installed are allowed to send app requests to her friends. Here, \(\langle \text{requester}, \text{install-friend} \rangle\) identifies users who are friends of the application user (i.e., Alice). The last one regulates whether applications her friends installed can send her notification of app requests. In this case, \(\langle \text{target user}, \text{friend-install} \rangle\) identifies the applications installed by Alice’s friends. The last two policies both deny the external com-
component of the applications to identify users or access users’ information who did not have them installed. Instead, the M1 or M2 modules can fulfill the task of sending out such requests without letting the external part learn the identities of those recipients.

Example 2: Access User’s Profile Information. Among the data that users put or generate in OSNs, some contains private information about users and should be safeguarded within the trusted OSN environment, while others may be appropriate for less trustworthy parties to access. For example, Bob prefers not to release his date of birth, relationship status, and email address to game applications he or his friends installed. On the other hand, the game application requires his age for customer survey, and normal key strokes and mouse moves during playing for proper functionality. To balance privacy and functionality, he expresses his preference through the following policies:

- Policy about Bob’s date of birth:
  \(<\text{access, dateofbirth, (owner, install)}, \{M1, M2\}>\)

- Policy about Bob’s key strokes:
  \(<\text{access, keystrokes, (owner, install)}, \{\text{external}\}>\)

- Policy about Bob’s email address for applications his friends installed:
  \(<\text{access, emailaddress, (owner, friend-install)}, \{M1, M2, M3, M4\}>\)

The policy about Bob’s date of birth allows the M1 and M2 modules of the application to access the data but denies any request from other internal modules and the external component. Key strokes in the game contains information that may need to be processed (possibly with significant computation) that is not likely to be done within OSN systems but more likely to be done at TPA, therefore is allowed to be accessed by the external component of the application. Similarly, the third policy indicates that only the M1-M4 modules of the applications that Bob’s friends installed are allowed to access Bob’s email address. This policy is used together with his friend’s policy when an application his friend installed requests to access his email address. Since there is no overt flow of private information between the internal modules and external components in the proposed framework, email address that the internal component accesses is not going to be transmitted to the external component.\(^2\)

5. CONCLUSIONS

We presented an access control framework for social networking platforms, preventing users’ private information from leaking to external parties. Our design splits third-party applications into internal and external components, allowing the internal components to access private information but keeping it away from the external ones. We provided a simple policy model for application-to-user policies to regulate application’s access. Users can specify different policies for different components of the same application, enabling more flexible and finer-grained control. Though the proposed approach does not eliminate privacy issues completely, it offers users greater controllability for their privacy against TPAs while allowing necessary features for the applications.

\(^2\)While these policies could be used to control an application’s access to user data, in real world system, the similar policies as a whole can be used to evaluate an application’s usages on a user’s private data at the time of the user’s initial access request to try out the application. If the data usages of the application modules satisfy the user’s policies, the user’s access to the application could be allowed. Otherwise, OSN may deny the user’s request or warn the users while showing which policies are violated by the application.

Acknowledgement

This research is partially supported by grants from the National Science Foundation and the State of Texas Emerging Technology Fund.

6. REFERENCES