Characterizing Privacy in Online Social Networks

Balachander Krishnamurthy
AT&T Labs – Research
Florham Park, NJ USA
bala@research.att.com

Craig E. Wills
Worcester Polytechnic Institute
Worcester, MA USA
cew@cs.wpi.edu

ABSTRACT

Online social networks (OSNs) with half a billion users have dramatically raised concerns on privacy leakage. Users, often willingly, share personal identifying information about themselves, but do not have a clear idea of who accesses their private information or what portion of it really needs to be accessed. In this study we examine popular OSNs from a viewpoint of characterizing potential privacy leakage. Our study identifies what bits of information are currently being shared, how widely, and what users can do to prevent such sharing. We also examine the role of third-party sites that track OSN users and compare with privacy leakage on popular traditional Web sites. Our long term goal is to identify the narrow set of private information that users really need to share to accomplish specific interactions on OSNs.

Categories and Subject Descriptors

C.2 [**Computer-Communication Networks**]: Network Protocols—*applications*

General Terms

Measurement

Keywords

Online Social Networks, Privacy

1. INTRODUCTION

Privacy leakage has been examined from various angles: characterization [11, 12], prevention techniques and tools [16, 1]. With the radical shift in the number of users worldwide onto online social networks (OSNs), there are new and significantly higher privacy leakage concerns as compared to traditional Web sites. OSN users are encouraged to share a variety of personal identity-related information, including physical, mental, cultural, and social attributes. Users who do this often believe that such information accessible to the OSN and maybe their "friends" on that OSN. In reality, the set of entities that can have access to various bits of private information is large and diverse: third-party advertisers and data aggregators, members in the OSN who are not friends of the user, and external applications. Also, if external actions taken by users while logged in to an OSN are tracked, such information can be used not just for marketing purposes, but shared with friends of the user leading to personal embarrassment. Facebook beacons[1] recently exemplified this problem and Facebook changed their policy.

For this work, we define the notion of privacy "bits" (pieces of information) for a user within an OSN[2], grouped together for setting of privacy controls, as shown in Figure 1. Ideally such grouping can be controlled by the user (as shown for the two users in the figure), but for existing OSNs such grouping is fixed by the OSN. Within an OSN, user's privacy settings dictate a left to right partial ordering of these groups with more private bits on the left and less private ones as we move right. Stacked privacy groups indicate groups with the same privacy settings. We explore this partial ordering in more detail as we examine the default privacy settings for OSNs as well as how OSN users make use of these settings.

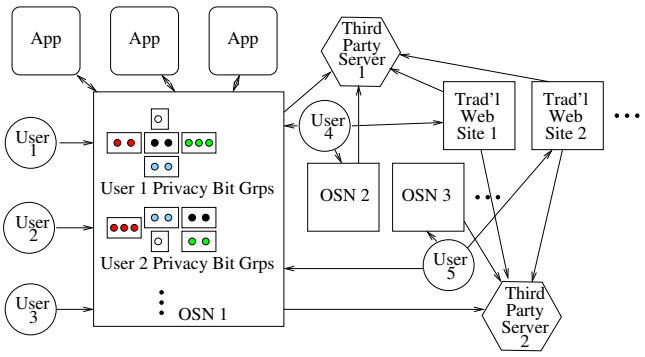

Figure 1: Privacy Information and Potential Leakage

Figure 1 also shows the flow of information among OSNs, external applications, third-party servers, and traditional Web sites. One current problem is that the manner in which

[1] `http://blogs.forrester.com/charleneli/2007/11/close-encounter.html`

[2] We note that there is no way to vouch for the accuracy of these bits such as friendship, relationship status, age, date of birth or even if the user is a real person.

WOSN'08, August 18, 2008, Seattle, Washington, USA.

private information is gathered by the various entities is often hidden. It is difficult for the user to know and control the various entities who can gain access to their information and to limit in such a way that it does not erode their ability to take full advantage of the various features of the OSN.

On some OSNs, such as Facebook, the choice varies from fine-grained settings for some features, to all or nothing in others. For example, when an external application is added by a user, the user *must* grant access to all of one's information to use the application. This is true even when the external application may need only a few relevant bits of private information or none at all. The granularity with which consent to access private information can be withheld is an important notion that we examine. Our goal is to see how to reduce privacy leakage while still enabling access to all the features of a OSN including use of external applications.

We also examine how OSNs group and provide a range of privacy settings for these bits. We study the default privacy settings for a variety of popular OSNs and if/how users modify them. The two dominant OSNs, Facebook and MySpace, account for a large portion of the market share [8, 13]. Other popular OSNs we examine include Bebo, Digg, Friendster, Hi5, Imeem, LiveJournal, Orkut, Xanga, as well as the smaller and newer micro-content network Twitter.

For these OSNs we characterize and study privacy, along with its potential for leakage by examining the bits of private information that a user supplies. We look at the allowable and default options for controlling what other OSN parties that can access the private information. We also study the involvement of third-party advertisers and data aggregators in tracking the actions of a user interacting with a OSN.

Section 2 describes the various bits of private information associated with user accounts on OSNs and who can access them. Section 3 examines if users change their privacy settings when allowed. Section 4 studies the role of third-party domains in aggregating user-related data and contrast it with their role in traditional Web sites. Section 5 looks at how privacy protection could be provided in OSNs by tailoring the actual privacy bits needed for specific interactions on OSNs. Section 6 presents related work and the concluding Section 7 includes future work.

2. USER PRIVACY CONTROLS

Figure 1 shows different bits of privacy information of varying types associated with user accounts. While terminology varies across OSN sites, such privacy bits are generally grouped to set privacy controls as follows:

Thumbnail A brief profile containing privacy bits with at least a user name (first and at times last name), and photo. A thumbnail is often the least private user information in an OSN.

Greater Profile This group has additional information: interests, relationships and other bits that a user is willing to provide. Protection settings for these additional bits of a user profile typically are through a separate privacy setting.

List of Friends The list of friends of a user where friendship requires agreement by the befriended user. Although a specific piece of information, this is often its own group with its own privacy setting.

User Generated Content Content added by the user such as photos, videos, comments and links. Not all OSNs provide separate privacy settings for controlling this group.

Comments Status updates, comments, testimonials and tags about the user or user content.

Although we have enumerated five groups above, it is conceivable that these may be teased apart into additional groups or merged into fewer groups. Coarser the granularity the harder it is to provide tailored privacy control.

OSN sites typically grant privileges to view these groups of information to three entities: the user, the user's friends or to all users. For example, for User 1 in Figure 1 the privacy groups from left to right could correspond to a group of bits only viewable by the user, followed by three groups viewable to the user's friends and a group viewable by all users. Some sites define an additional access level that includes more users than friends, but not all users of the site. For example, Facebook allows viewing privileges to be granted to the members of a user's networks, which include geographic regions, schools, and work. MySpace allows viewing to be granted for users 18 years or older. Bebo allows viewing of information for a user-defined age range. Orkut, Friendster and Facebook (recently added) allow viewing privileges to be assigned to friends of friends. Separate from these privileges assigned to a party of users, many OSNs also allow them to be *denied* to specific users.

As an example, Table 1 shows the allowable privacy settings for viewing different privacy bit groups in Facebook and MySpace. We study how these settings are used in practice in Section 3. A "○" indicates that it is possible to allow the given party to view the information in the privacy bit group while a "-" is used to indicate that it is *not* possible. So Table 1 shows the additional privacy bits that are part of a Facebook user's Greater Profile (referred to simply as *Profile* in Facebook) can be set to be viewable by the user's friends, friends of friends, or set to the user's friends and users in the same network. Note that user generated content is also included in this group for Facebook. It is not possible to grant this privilege to all users nor restrict it to only oneself. A box around an option (e.g. ⊡) indicates the default value for a setting. Thus, "friends+networks" is the default Facebook setting for viewing a user's greater profile.

Table 1: OSN Privacy Settings for Viewable Information

Facebook					
Privacy Bit Group	Self	Friends	Friends of Friends	Friends+ Networks	All
Thumbnail	-	○	○	○	⊡
Greater Profile	-	○	○	⊡	-
List of Friends	-	○	○	○	⊡
User Gen. Content	-	○	○	⊡	-
Comments	○	○	○	⊡	-
MySpace					
Privacy Bit Group	Self	Friends	Friends of Friends	Friends+ Age>18	All
Thumbnail	-	-	-	-	⊡
Greater Profile	-	○	-	○	⊡
List of Friends	-	○	-	○	⊡
User Gen. Content	-	○	-	○	⊡
Comments	-	○	-	○	⊡

The privacy settings in Table 1 show that the *networks* concept is important to privacy control for Facebook. By default, information in a user's profile/content, and comments (as on a user's "Wall") are viewable by any other user in the user's networks. Facebook also allows a user to designate a primary network and limit access to information for friends and users in the primary network. Facebook provides some controls on who can join a school network (via age claims and needing to be befriended by someone in that school network), college and work networks (via email address in those domains), but there is no control on who may join a regional network. In addition, the thumbnail profile and list of friends for all Facebook users is publicly available to all other Facebook users. As shown in Table 1 all of these settings can be changed via the Facebook interface, but the default settings allow any user to gain access to all information about another user on Facebook. Columns to the right of the double vertical line are particularly problematic for privacy as these are situations where a user is not able to control which users are able to view the given information. The only exception is that Facebook imposes age restrictions that are not explicitly documented on their Web site, but we have observed it and it has been reported [15]. These restrictions prohibit users over 18 years of age from viewing profiles of users under age 18 and vice-versa, unless the users are friends. However, these restrictions do not hide a user's list of friends from other users across this age divide.

Table 1 shows that MySpace uses similar permissive defaults in terms of access to a user's information—all users have access to all other user's information. There is no way to even limit the visibility of a user's thumbnail profile. MySpace uses coarse-grained privacy controls where a single setting controls access rights for all other rows in Table 1—in terms of Figure 1 all privacy bits other than the thumbnail are combined in the same privacy bit group. The range and coarseness of privacy controls for Bebo are the same as MySpace except that instead of fixing the age limit of 18, Bebo users can specify an age range to use as a level for granting access. In addition, Friendster, Hi5, and Xanga are similar except they do not have an extra access level. Imeem and LiveJournal allow a user's profile to always be viewable by all users. Twitter has a single privacy setting to control which decides who can see their status updates, location, or biographical information. Digg allows viewing of some bits in the profile and comments to a user to be controlled. Orkut actually restricts the default settings for some parts of the greater profile to be only viewable by friends.

In summary, the privacy groups for OSNs are large with many bits controlled with a single setting, OSNs do not provide any range of privacy settings for some of the privacy groups, and all OSNs have permissive default settings that allow viewing privileges to more users than just friends. The end result is that by default a user does not control who has access to their information on these sites unless they explicitly control their privacy settings.

3. USER PRIVACY SETTINGS

The default privacy settings for OSNs are permissive in allowing strangers in an OSN to access user's information. We now examine the extent to which users change their privacy to more restrictive settings. A 2005 study found that only 1.2% of college Facebook users at CMU changed the searchability of their thumbnail profile (first row in Table 1) and only 0.06% changed their profile visibility (second row) [7]. More recently, it was reported that 75% of 200 users in the Facebook London regional network have their full profile viewable by other users in the network [14].

We studied privacy settings for Facebook and MySpace, the two dominant OSNs, as well as Bebo and Twitter for which obtaining settings was available to us. A crawl carried out for a different study [9] of over 67000 Twitter users found that more than 99% had retained the default privacy setting whereupon their name, list of followers, location, Web site, and biographical information are visible.

For MySpace, we examined the percentage of users that allowed their profile to be viewable—the default setting. A user in MySpace is assigned a numeric userid that is used for viewing their profile (both the thumbnail and greater portion). We generated 5000 random numeric userids in an observed range of valid userids and in February 2008 retrieved their corresponding user profiles. We obtained profile information for 3851 valid userids, of which 79% (3046) of users retained their default setting that their profile, friends, comments and user content were viewable.

We studied Bebo likewise, but instead of random userids we examined the profiles of users who were members of interest groups within Bebo. Again in February 2008, we found 80% of the Bebo users we examined allowed their profile, friends, comments and user content to be viewable.

We took a different approach to study Facebook, using its 506 regional networks (circa April 2008) that represent geographical areas. We did this because Facebook restricts public profile viewing to users in the same network and there are fewer controls on who can join a regional network, although a user can only be a member of one regional network at a given time. In the U.S., the 272 regional networks correspond to cities but often they include users who may live nearby. Outside the U.S., the 234 regional networks correspond to cities in Canada and U.K., but to countries elsewhere. Table 2 shows the 20 U.S. and Table 3 the 18 non-U.S. regional networks studied during April 2008.

Our choice of regional networks was done to meet three criteria: first we wanted both large and small regional networks in terms of number of users; second we sought some degree of geographical diversity in the U.S.; and third we tried for linguistic and cultural diversity in the non-U.S. regions. We began with a four-part size-based separation and then chose within each size range cities (for U.S.) and countries (for non-U.S) balanced by size as well as geographic (for U.S.) and linguistic/cultural diversity (for non-U.S.). Although our sample may not be truly representative of the global community of Facebook users, we believe it represents a broad cross-section of regional networks.

We used the random network browsing feature of Facebook which returns thumbnails for up to ten random users on each retrieval. All retrievals were made by an over 18 user who is a member of the regional network being tested. Along with the thumbnail for each user, the search returns HTML links to view the profile and view friends of the user, if the user's privacy setting allowed it. Based on default settings, viewing the profile is allowed for all users in the same network while viewing of friends is allowed for all Facebook users. We made 200 successive retrievals for each regional network; up to ten users are returned each time. Users who have changed their thumbnail privacy setting to not be viewable by their network will not be found by this approach,

but [7] found only 1.2% had changed this setting. Due to the repeated random sampling duplicates occurred and were eliminated in the analysis. We also eliminated cases where viewing profiles of users apparently under 18 (surmised if a high school student had not yet graduated) were disallowed because of Facebook's restriction.

Our 20 U.S. regional networks and their privacy settings results are listed in Table 2 in decreasing order of the number of network users as reported by Facebook. All results represent at least 1600-1700 users after elimination of duplicate entries. This sample size results in a 2-2.5% error margin with a 95% confidence level.

Table 2: Privacy Settings in U.S. Facebook Regional Networks

Regional Network	Users (K)	%View Profile	%View Friends	Regional Network	Users (K)	%View Profile	%View Friends
New York,NY	866	53	78	Syracuse,NY	54	75	90
Chicago,IL	649	54	78	Worcester,MA	45	77	94
Los Angeles,CA	595	62	82	Peoria,IL	44	77	93
Atlanta,GA	390	56	82	Boise,ID	36	83	96
Dallas/FW,TX	336	63	84	Tupelo,MS	29	76	98
Seattle,WA	210	64	83	La Crosse,WI	25	71	94
Sacramento,CA	99	76	90	Monroe,LA	21	79	98
Des Moines,IA	83	67	85	Ithaca,NY	17	78	95
Okla City,OK	80	71	87	Abilene,TX	10	82	97
Greenville,SC	66	72	90	Casper,WY	6	84	99

The results in Table 2 show a negative correlation between network size and percentage of users with viewable profiles ranging from 53% for Chicago to 84% for Casper with a strong linear correlation coefficient of $r = -0.88$. Similarly the percentage of users who allow their friends to be viewed is even higher and also shows a negative correlation with network size at $r = -0.85$. We do not know the reason for this strong negative correlation without surveying user attitudes, but hypothesize that users in smaller networks are less concerned in making private information available.

We used the same methodology to study results for non-U.S.-based regional networks. The results, shown in Table 3, show a similar negative correlation with network size as found in the U.S. results, with $r = -0.80$ for viewable profiles and $r = -0.86$ for viewable friends. However it is interesting that the correlation still holds across the many cultures that these networks represent. The London value of 51% of users with a viewable profile is lower than what had been reported in a smaller study last year [14].

Table 3: Privacy Settings in Non-U.S. Facebook Regional Networks

Regional Network	Users (K)	%View Profile	%View Friends	Regional Network	Users (K)	%View Profile	%View Friends
London	2486	51	76	Brazil	118	87	96
Australia	2015	63	83	Edinburgh	98	75	93
Turkey	1866	50	76	South Korea	71	79	88
South Africa	646	65	88	Jamaica	41	72	91
India	633	68	86	Iceland	28	84	97
Hong Kong	520	59	82	Iran	21	91	97
Mexico	448	73	90	Algeria	10	92	98
Singapore	382	70	88	Angola	2	91	98
Greece	241	70	91	Nauru	0.2	93	96

The results in both tables show that users appear to place a higher value on the privacy of their profile information compared to their list of friends. We say apparently because these two settings are not set in the same place within the Facebook user interface and so differences may be due to interface issues. As shown in Table 1, another privacy control setting in Facebook determines who is allowed to view a user's Wall of comments. Instead of being available as a separate setting, it is a subset of the View Profile setting—so it can only be applied as a further restriction. We examined this setting for a portion of the U.S.-based regional networks in our study. In the New York region we found 79% of those with a viewable profile (42% of all users) allowed their Wall to be viewable to anyone in the network. It was 83% (53%) for Seattle and 95% (73%) for the Worcester region. A user's Wall is thus the most protected privacy bit in Facebook, although this is necessarily so by its inclusion within the View Profile setting for a user.

In summary our results show that while there is now some use of privacy settings by OSN users, there is still a significant portion of OSN users who have not changed their permissive settings and allow unknown users to view private bits of information. We did find a consistent negative correlation between the use of privacy settings and network size across both US and non-US-based regional networks.

4. USE OF THIRD-PARTY DOMAINS

Beyond revealing private information to other users within an OSN, another potential source of privacy leakage is the tracking of user actions by third-party advertisers and data aggregators. In prior work, we defined and measured the privacy footprint for a collection of roughly 1000 popular traditional Web sites [11]. This privacy footprint is the set of interactions formed when user retrievals of first-party Web pages cause retrievals of objects from, often hidden, third-party domains[3] as shown in Figure 1. These third-party domains often act as aggregators of a user's traversals on the Web. We found a large number of associations between first-party sites via shared third-party aggregation domains.

A natural follow-on question to this previous work is to ask to what extent these third-party domains are used by OSNs. These domains are particularly troubling for privacy with OSNs because unlike most Web sites, users login to OSN sites and store personal information about themselves on these sites. If OSN sites are also making use of third-party domains that are tracking user visits to these and other Web sites then there is an even greater potential for privacy loss. Previous work does conjecture that third-party aggregation will be used less for OSNs because they encourage users to "live" on their site [3].

In order to measure the use of third-party domains, we first established a set of actions to take for each of the OSNs in our study. These actions are not exhaustive for any of the OSNs, but represent common actions of a user while interacting with each OSN site—whether they be interacting with friends on Facebook or viewing a comment made by a user on Digg. The specific set of actions used for each OSN site, which we refer to as a *session*, is:

1. Login to site with account/password.
2. View friends or user contributed content.
3. Look at a friend's or contributor's profile.
4. Send friend a message or comment on content.

[3]Servers sharing the same 2nd-level DNS domain are grouped within the same domain.

Table 4: Top Third-Party Domains Used by OSN Sessions

	Online Social Network										
Third-Party Domain	Bebo	Digg	Facebook	Fr'ster	Hi5	Imeem	LiveJ	MySpace	Orkut	Twitter	Xanga
doubleclick.net	✓	✓		✓	✓	✓	✓	✓			✓
2mdn.net	✓	✓		✓	✓	✓		✓			✓
advertising.com	✓		✓	✓	✓	✓		✓			
atdmt.com		✓	✓	✓		✓	✓	✓			
googlesyndication.com	✓			✓	✓		✓	✓			
quantserve.com		✓			✓	✓					✓
adbrite.com	✓			✓		✓					
google-analytics.com				✓		✓				✓	✓
yieldmanager.com	✓				✓						✓

5. Return to user home.
6. View networks/groups.
7. Look at members in a network/group.
8. Logout from site.

While this session was being executed at each OSN, the set of retrieved objects were recorded via the "Pagestats" Firefox browser-extension [5], which records information about each HTTP request and response. A session was performed manually for each of the eleven OSNs with Pagestats logging the retrieved objects. The post-processed log yielded the list of third-party domains where at least one object was retrieved for each OSN. This session and analysis was repeated five times for each OSN after observing that not all actions always caused the same set of third-party objects to be retrieved. Table 4 shows key results of our analysis with the third-party domains most widely used. A '✓' is used to indicate situations where the third-party domain was used in the majority of the five sessions executed at the given OSN. The 9 third-party domains with at least three ✓'s are shown in the table. Another 12 third-party domains (not shown) occur in a majority of sessions for two of the OSNs and another 36 third-party domains occur in a majority of sessions for one of the OSNs. To understand the sensitivity of our results to the session of actions we choose, we extended our session for a subset of the OSNs to include writing on a comment wall, using an application, joining a group and editing the user profile. These additional actions did not result in the use of additional third-party domains except for the applications, which caused the browser to go to new third-party domains used by these applications. We observed approximately one new third-party domain accessed for each new applications we used.

Table 4 indicates much common usage across OSNs with `doubleclick.net` and `2mdn.net` consistently represented across the 11 OSNs. Overall there is high usage of these third-party domains by most OSNs. The median number of unique third-party domains contacted at least once for an OSN was 25 and the median number of unique third-party domains contacted at least once per OSN session was 12. Even Orkut, which is shown using none of the top third-party domains in Table 4 is part of the Google network of domains.

To understand the significance of these third-party domains in making associations not only between OSNs, but also with traditional Web sites, we re-ran our study of roughly 1000 popular URLs from [11] in February 2008. We used the batch-mode feature of Pagestats and show results from 967 URLs. The third-party domain results of this updated study for popular Web sites show that 30% of these popular sites are now using `google-analytics.com` (vs. 7% in [11]). Overall the results represent over a 30% increase in the penetration of the top-10 third-party domains amongst popular Web sites relative to April 2006 results in [11]. Particularly noteworthy for this work is that the third-party domains listed in Table 4 are *all* in the top-10 third-party domains for popular Web sites.

The use of third-party domains with the capability to track user activity is pervasive for OSNs even as it continues to grow for traditional Web sites. This trend is of particular concern as OSN users being tracked have explicitly identified themselves by logging into the OSN and provided the OSN with personal information.

5. PRIVACY PROTECTION

Users are generally unaware of who has access to their private information on OSNs. Interestingly, most users may be able to carry out a large fraction of their actions on OSNs while significantly shrinking the amount of private information that is made available to others. Most of the thousands of popular applications on OSNs like Facebook do not need complete access to the private information of users, yet Facebook gives users no choice if they want to download and use an externally created application. For example, it is hard to justify the use of information other than the list of friends is required to run popular applications such as Scrabulous.

We need a way to enumerate the precise private bits of information that are *actually* needed for a user to interact with and make full use of the myriad features of a OSN. While the privacy bits may vary with a specific feature (some external applications may genuinely need more information than others), we can do much better than the current all or nothing approach. Limiting access to just friends or those in a network is not fine-grained enough. We could start bottom-up and hand out information to more and more users/networks based on need. Just as there are groups, networks, etc. as aggregating mechanisms there should be a way to *deny* private information at each aggregation. It should be possible to both deny and enable access to private information at the same level of granularity.

OSNs must clearly indicate the *bare minimum* of private information needed for a particular set of interactions. If an external application requires access to list of friends and nothing else, then the default should be that bare minimum. If additional features of the application require access to other bits of private information then access to the supre-

mum of the information could be enabled, and no more.

A mechanism to identify the metrics *bare minimum* and *supremum* would be a useful addition to the privacy arsenal. Such metrics would allow us to compare various OSNs on an equal footing and let the users decide how comfortable they are with the privacy information that is being shared. Our enumeration of privacy groups in Section 2 and the left-to-right ordering of privacy groups in Figure 1 is a first step in this direction. A user could have a threshold mark along this spectrum in terms of what groups (and thus what bits) they are willing to share freely. For each set of interactions or use of an application, the OSN could indicate what bits are needed and if they are to the right of the user's threshold, access is provided transparently. If some additional bits to the left of the user's threshold are essential then the user can be prompted. The user can allow or disallow, and optionally set the duration (for current session, forever etc.) for such a grant. There are several browser extensions that are already capable of providing functionality similar to this in other, related arenas such as security and script execution privileges. It would be relatively easy to create an extension that can be used with OSNs.

6. RELATED WORK

There has been considerable work in the privacy and anonymization field on traditional Web sites related issues. There is a growing amount of literature [2, 6, 17] on OSNs as well, although the focus is on examining identity leakage due to attacks or data being published. We focus on the scenario where no additional information is actively released about social network data and with no explicit adversary. However, the work on re-identification [4] that allows one to relate supposedly anonymous data with actual identities by combining external data is related. In OSNs it is easier to get various bits of private information and each of them can be used to merge with external information to identify a person. Given the recent large-scale attempts to have medical records online, such concerns are well-motivated.

Privacy protection measures exist for Web browsing in the form of well known extensions in Firefox, anonymized access etc. However, in the presence of voluntary information provided by users, what matters most is *limiting* the access. Strong co-operation is needed by the OSN and their guiding policies will be a key factor in protecting user privacy. The recent withdrawal by Facebook of default opt-in to *beacons* whereby visits by users of certain third party e-commerce sites could trigger automatic notification to their OSN friends, is a step in the right direction.

7. CONCLUSIONS

We characterized and measured various privacy aspects across eleven OSNs. Users willingly provide personal information without a clear idea of who has access to it or how it might be used. The range of privacy settings that OSNs provide were found to be permissive since default settings allow access to strangers in all OSNs. We studied how users make use of privacy controls to limit access and found that between 55 and 90% of users in OSNs still allow their profile information to be viewable and 80 to 97% of users allow their set of friends to be viewed. We found a strong negative correlation between regional network size in Facebook and the use of these privacy settings to limit access. Much like traditional Web sites, third-party domains track user activity pervasively in OSNs.

Various extensions are now feasible. We proposed a new idea to better match what information a user makes available with what is needed by other users and applications. While focusing on regional Facebook networks, we also did preliminary investigation of a college (5900 users) and a high school network (2200 users) to which we had access and found 80% of users allowed their profile to be viewed for these much smaller networks. Techniques to protect privacy leakage to third-party domains for traditional Web sites was investigated in [10] and could be extended to OSNs. Finally looking at privacy implications for interactions between OSNs, such as between Facebook and Twitter, is another direction of future work.

8. REFERENCES

[1] Anonymizer–online privacy and security. `http://www.anonymizer.com`.

[2] Lars Backstrom, Dan Huttenlocher, and Jon Kleinberg. Wherefore art thou R3579X? In *Proceedings of the WWW*, 2007.

[3] Graham Cormode and Balachander Krishnamurthy. Key differences between Web1.0 and Web2.0. *First Monday*, 13(6), June 2008.

[4] Mark Crovella and Balachander Krishnamurthy. *Internet Measurement: Infrastructure, Traffic and Applications*. John Wiley and Sons, Inc., 2006.

[5] Scot DeDeo. Pagestats, May 2006. `http://www.cs.wpi.edu/~cew/pagestats/`.

[6] Korolova et al. Link privacy in social networks. ICDE 2008 Poster. `http://www.stanford.edu/~korolova/link_privacy_ICDE08.pdf`.

[7] Ralph Gross and Alessandro Acquisi. Information revelation and privacy in online social networks (the Facebook case). In *Proceedings of the ACM Workshop on Privacy in the Electronic Society*, November 2005.

[8] Social networking visits increase 11.5 percent from January to February, March 2007. `http://www.hitwise.com/press-center/hitwiseHS2004/socialnetworkingmarch07.php`.

[9] Balachander Krishnamurthy, Phillipa Gill, and Martin Arlitt. A few chirps about Twitter. In *ACM SIGCOMM Workshop on Online Social Networks*, August 2008.

[10] Balachander Krishnamurthy, Delfina Malandrino, and Craig E. Wills. Measuring privacy loss and the impact of privacy protection in web browsing. In *Proceedings of the Symposium on Usable Privacy and Security*, pages 52–63, Pittsburgh, PA USA, July 2007.

[11] Balachander Krishnamurthy and Craig E. Wills. Generating a privacy footprint on the Internet. In *Proceedings of IMC*, October 2006.

[12] Bradley Malin. Betrayed by my shadow: Learning data identify via trail matching. *Journal of Privacy Technology*, June 2005.

[13] Social network downtime in 2008, February 2008. `http://royal.pingdom.com/?p=253`.

[14] Facebook members bare all on networks, Sophos warns of new privacy concerns, October 2, 2007. `http://www.sophos.com/pressoffice/news/articles/2007/10/facebook-network.html`.

[15] Brad Stone. New scrutiny for Facebook over predators, July 30, 2007. `http://nytimes.com/2007/07/30/business/media/30facebook.html`.

[16] Latanya Sweeney. k-anonymity: a model for protecting privacy. *International Journal of Uncertain. Fuzziness and Knowledge-Based Systems*, 10(5):557–570, 2002.

[17] Bin Zhou and Jian Pei. Preserving privacy in social networks against neighborhood attacks. In *ICDE*, 2008.