Privacy and Online Social Networks: Can colorless green ideas sleep furiously?

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

ABSTRACT

One definition of privacy is selective revelation of information about oneself. In the age of a billion user social networking world, it has become increasingly difficult for people to control what they are disclosing to whom. Current privacy protection measures block leakages via privacy settings that are *syntactic* in nature. The title borrows an example from Chomsky[1] who used it to provide a sharp distinction between syntax and semantics. Virtually all privacy solutions thus far attempted handle issues relating only to the first hop of the personal data flow from the user. Existing solutions do not attempt to cover all the entities who might end up receiving the data, ensure the need for or use of the data collected, the duration of data retention, or if the data could be merged with external information to reveal the user's full identity. The gap can only be filled by examining the semantics behind the multi-hop flow of user's data over time. This paper surveys the state of the art and present some potential directions in moving from a syntactic approach to a more holistic semantics-based approach.

1. INTRODUCTION

Online Social Networks (OSNs) continue to grow in terms of number of users (roughly one of two people who have access to the Internet have an account on an OSN), network traffic, user time, and correspondingly commerce in the form of advertisements. Targeted advertisements have been touted as the holy grail requiring tailoring of ads to the interests inferred based on users' actions on OSNs or other websites. The profile information provided by the users to the OSN, their social graph, their set of interactions via internal and external applications, linkable actions outside the OSN, are all input to generating and delivering targeted ads. Clicks on the ads flesh out the picture and over time the longitudinal data gathered ends up reflecting the value of the user to the OSN.

Privacy revolves around the notion of selective revelation of information about oneself. Increasingly, users disclose information via multiple input vectors (mobile devices, laptops, PCs) to numerous entities, some of whom are invisible to the users. Personal information disclosed include location, device identifiers (an IP address or unique ID of a cell phone), contact (email and physical addresses, telephone numbers), identity (name, date of birth, photos), social (friends, interests), and activities (search, Web site visits, games) among others. Further, data collected tends to last with the added risk of being linked in diverse ways. The combination of disclosure, storage, and linkage is the core of the privacy problem faced by users.

Most users derive significant value through memberships in OSNs and interactions with popular Websites, while the underlying economics of the transaction remains hidden. Privacy activists and consumer protection agencies have been working to resolve the tussle with the advertising industry. The Web Consortium's Do Not Track (`http://donottrack.us`) effort envisions a simple syntactic specification of a user's intention not to be tracked. Much of the discussion has been about the disposition of data collected. However, there are continuing battles over the meaning of exactly what tracking denotes. Given the size of the online ad industry (valued at over $60 billion) and a huge ecosystem supporting this effort, the incentives to continue tracking are enormous with little potential for abatement. There is little agreement on the nature of the actual problem or understanding of the diverse threat models. The market for user's data is not transparent.

Virtually all of the current privacy protection mechanisms operate at a *syntactic* level—individual personal information bits (such as date of birth, search strings, cell phone numbers) could be protected from being shared with the party with whom the user is currently communicating. Aspects of the privacy problems include use of data across subsequent communications with multiple parties; this is the "secondary use" [17] problem. Some may be hidden, so there is a strong need to examine the operational semantics arising from the full information flow. The data is also used over a long period of time and across different OSNs (data aggregation) and in an entirely different context compared to the one in which it was shared (situational semantics where privacy boundaries differ). Without the user's knowledge, the data is linked with other auxiliary information [2]. The millions of external applications available to OSN users are written and hosted by entities other than OSN. At installation time, the applications request for access to various privacy bits. It is impossible for an OSN to verify if all these bits are actually required by hand. Typically OSNs allow the users to decide if they want to share the bits or choose not to install the application.

The title of the paper refers to an example that Chomsky used to crisply capture the key difference between syntax and semantics: the sentence is syntactically correct but has no real meaning. Likewise, most of the current privacy pro-

[1] `http://en.wikipedia.org/wiki/Colorless_green_ideas_sleep_furiously`

[2] http://research.microsoft.com/pubs/64346/dwork.pdf

tection attempts are heavily biased towards the syntactic and do not solve the privacy problem. Tracking the full semantics facilitates in understanding both when and what context the information is shared. This leads to better potential privacy control. A semantic approach to the privacy problem would examine the full flow of user's data, all the parties with whom data might be shared, the set of privacy bits actually needed for the successful operation of external applications, and the time and context in which the shared data might be used. A user purchasing jewelry for a significant other may not want that information to be shared with all their friends as it might include the significant other. Limiting disclosure to certain contexts would reduce the risk of such an occurrence.

This paper traces the evolution of and complexity associated with the privacy problem. Given both the increasing role of external applications and linkages with online and offline data, current syntactic protection methods are simply inadequate. Section 2 looks at the various vectors of privacy leakages and potential linking of the leaked data elements optionally with other ambient data available. Section 3 presents a taxonomy of the current range of proposed solutions, which are by and large syntactic. Section 4 explores what can be gleaned from gains in the related security arena and examines the role of economics. Section 5 presents concrete examples of the need for semantic analysis and the shortcomings of the current syntactic approaches. Section 6 proposes a way to move towards including semantics in privacy protection.

2. LEAKAGE AND LINKAGE

The range and diversity of interactions between users via their social graph, across multiple OSNs and the Web, and the external applications ecosystem, have expanded dramatically. The concerns of privacy leakage has grown alongside. Users have to manage their privacy without fully understanding the breadth of the problem. The complexity of potential flows of information and the consequences of actions over time is too large to envisage a coherent protection mechanism. Here, we trace the discoveries of leakage over multiple axes: across time, passive leakage via regular OSNs and their mobile counterparts, and inference through active mining. We then examine the potential for linkage of data.

An early look [13] at privacy settings and availability of information within OSNs showed that data about a large number of users was available to significantly more people than might be expected. While default settings may have been the culprit early on, the sheer number of privacy settings and the complexity of tracking them over time was clearly overwhelming to most users who ended up relying on the OSN for the best default settings. Subsequent work [14] showed simple vectors of leakage enabled by interaction with OSNs. In some cases raw bits of personally identifiable information were being leaked directly. The OSN identifier was being leaked via HTTP headers as well as popular external applications. Some of these leakage vectors were plugged (often only after widespread publicity in the media[20]) and some triggered governmental actions that took a few years[3] to result in settlements. Meanwhile numerous other leakages were disclosed followed by quicker reactions from the industry and growing interest from the privacy activists and government agencies. With the explosion in mobile devices and their use in accessing OSNs, new leakages identified [15, 8] included current presence on the OSN, unique device identifiers, user locations etc.

Another vector for privacy loss is attributes that can be inferred through active mining. User profiles can be inferred [10] in OSNs by exploiting the sociological concept of homophily (the common tendency of humans to have more affinity for people with likes similar to their own) and thus identifying communities. Other mining techniques to reap email addresses by automated query[1] and their uniqueness and traceability [11] have shown the potential for linkage.

The ability of aggregators to merge publicly available personal information in the OSNs with external information is worrisome. The ease of linkage of user's data [12] means that, even in the absence of cookies, tracking has grown. Leakage of highly privacy-sensitive search strings (e.g., names of specific diseases) that can be linked to an user's OSN account via globally unique identifiers (such as email addresses) raises significant concern. A user's action in an OSN indicating their endorsement of a business provided an additional linkage mechanism for tracking outside the OSN.[4] The connections across OSNs showed the potential reach due to transitive closure of the flow of a user's data. For example, a user's "check-in" in a mobile OSN indicating their current location could be translated into a tweet on Twitter or status update on Facebook. Users may have explicitly enabled such a communication when they opened their accounts. Over time new linkages occur and new applications become available. Yet there is no mechanism for feedback to the users about these changes—or the extent to which their data can spread as a result (sometimes caused by a single action at a single OSN).

By merging online and offline data a much richer picture of users can be generated and sold to interested parties. For example, local courts have personal information about legal cases, deeds, criminal records that are not generally available online. Fairly recently information about real estate transactions and political donations of individuals have already migrated online. Combining personal information that is available on social networks with offline data leads to a significantly broader profile of the user. This cannot be prevented by any syntactic method of privacy protection. While an OSN cannot provide protection to users who already have a lot of external data available, they can ensure that they are not an unwitting vector for leakages by contributing to potential linkage. Instead, the OSNs should use their platform to educate and alert their users about vulnerabilities. Later we will see examples of technical means by which OSNs can offer such help.

3. ATTEMPTS TOWARDS A SOLUTION

Popular privacy protection proposals are almost always syntax-based. A brief taxonomy of attempted solutions to the privacy problem includes browser-level protection mechanisms, new architecture proposals, use of cryptography, and masking identities. Governmental consumer protection agencies and others have also made attempts at promoting collaborative efforts.

As browser-based solutions are easy to deploy and attract a large number of users, many of the protection attempts

[3]http://ftc.gov/opa/2012/05/myspace.shtm

[4]http://ssrn.com/abstract=1717563

have focused there. The attempts can be categorized as presenting detailed information to the users, such as PrivacyBucket (use demographic information to predict what can be learnt about users), visualization tools (e.g., WebCrumbs, PrivacyDashboard), FourthParty (instruments in-browser functions and logs all resource accesses and cookies into a searchable database), and Priv3 (enables Facebook's Like button only if the user affirmatively interacts with the site). Some popular browser extensions have been re-purposed towards protection; for example AdBlock and Ghostery can prevent connections to aggregators and NoScript can prevent execution of suspect JavaScript. To address the growth of aggregators, users can contribute regular expressions that block connections to the new ones to AdBlock's shared database. Globally unique identifiers constructed by stringing together version numbers of the extensions or font collections installed in a user's browser, negate protection provided by blocking all cookies and disabling JavaScript. All these syntactic protection are merely discrete steps in the absence of a comprehensive privacy solution.

Rather than tinker at the edges, entirely new alternatives to the currently centralized OSN model have been proposed at the architectural level. Architectural approaches bypass the syntax and semantics question by moving the users into an entirely new milieu where privacy can be addressed comprehensively. Projects like Vis-a-vis [19] and Safebook[3] partition the user's data and provide privacy as a collateral benefit. If the data is stored among an available set of peers the need for third parties and advertisements is eliminated. But architectural attempts lack a viable economic‘ model and fail to address two key issues: guaranteeing availability of all user's data at all times and the proposed system scaling up to attract a large number of users. None provide viable incentives to *move* users from existing OSNs.

Reusing security paradigms is often proposed for solving privacy problems. This includes cryptographic privacy for OSNs like Twitter[6] whereby user's tweets are protected to limit what various parties can learn and controlling one's identity cryptographically [7] while participating in electronic commerce. But unless the end-to-end semantics are completely protected, leakage is bound to occur.

The notion of fencing user's data and communications is another thread of research; this includes sending false or misleading information to a site. For example, consider a user who wants to take advantage of location based services but does not want to reveal her location. A user's device could send queries about multiple nearby locations and locally reconstruct the answer for its current location and thus avoid revealing her current location [18]. Such attempts fall under the rubric of trying to mask one's personal information (such as routing requests through proxies or using Tor) and by their nature are syntactic. But, as human rights activists—who really understand the risks they face—know that such approaches to protecting individual privacy bits (IP address, location information) need to be merged so the full collection of user's private data can be protected.

Another reason for failure of even the limited privacy protections available is their limited *usability*. Factoring in the additional semantic complexities currently ignored by protection mechanisms would worsen usability further. The large number of privacy settings and their complex interactions has made it hard to present a simple interface to end-users to manage their privacy. Users are often unclear about their privacy needs [4, 9] which worsens the situation.

There has been prior work on examining semantics of privacy in the areas of contextual integrity [16, 5] and accountability[5] —these have been at the theoretical level. A framework on usage control policies to improve compliance has been proposed [6] in the OSN context. Information tracing attempts at semantic level (e.g., TaintDroid[8] and Pios [7]) assume that the information is either in the users control or the application binary is available; neither are true in the OSN context.

Finally, we look at some non-technical efforts made to address the privacy problem. The U.S. Federal Trade Commission has run privacy roundtables with privacy advocates and ad industry personnel participating. The workshop on Web Privacy Measurement[8] brought privacy researchers and tool builders and representatives of governmental agencies of US, Canada, and Europe. The World Wide Web consortium has been working on the Do Not Track mechanisms with representatives of ad industry and publishers. The various OSNs, publishers, and the advertising community want to ensure viability of the online ad industry as popular Websites and OSNs rely on advertising rather than subscription fees.

4. SECURITY AND ECONOMICS VIS A VIS PRIVACY

We now look at what can be learned from progress in security over the years. We then discuss the increasing role of economics. OSN services are subsidized by advertisement, and bringing the hidden economic transaction to the forefront clarifies matters.

Compared to gains made in the security arena, there have been few advances in privacy protection thus far. Next, the adversary model is generally better understood in security than it is in privacy. Even advanced users and privacy advocates do not have a good handle on the entire range of privacy threats. It is virtually impossible for users on their own to track the full extent of information spread or even identify the entities who may receive it. Just as early Internet protocols were designed without consideration of security issues, the OSN ecosystem has evolved largely with privacy as an afterthought at best. Finally, while security is largely a binary property, privacy requirements of users can fall along a spectrum.

The security market is different than the privacy one. Many security products are available requiring minimal end-user configuration. Hundreds of millions use pre-configured firewalls with their home routers; modern operating systems come with anti-virus systems, and browsers are equipped with various security alerting mechanisms albeit with usability issues. In contrast, the default settings for privacy are not always the most desired ones. There are no software packages with widespread applicability that are pre-installed on devices or browsers. This could be due to either the perceived absence of a market for privacy products or current lack of agreement on a clear threat model.

There is, however, a large market for continued access to

[5]http://dig.csail.mit.edu/2011/Papers/IEEE-Policy-httpa/paper.pdf
[6]http://dig.csail.mit.edu/2010/Papers/Privacy2010/tkang-rmp/paper.pdf
[7]http://www.cs.ucsb.edu/ chris/research/doc/ndss11_pios.pdf
[8]http://www.law.berkeley.edu/12633.htm

user's data. Recalling the adage "if you are not a consumer then you are the product", users are better off understanding the hidden economic transactions.

As clearly enunciated by Acquisti in his early work on economics of privacy,[9] there is a need for the participation of the industry, government/law, and implementation by technologists of an acceptable economics-based solution. There have been advances in bringing the economics aspect of privacy to the fore but the problem remains unsolved.

An interesting evaluation[2] of privacy and economic motivation of OSNs pointed out that user decision making is not always rational. A small field study[10] showed that users would give more personal information for a small discount on the price of DVDs; increased privacy protection did not lead to better product sales although the users in a post-study interview claimed to care about privacy!

5. THE NEED FOR SEMANTICS

So far we have examined vectors of leakage and linkage, the proposed solutions, and the lessons that could be drawn from security and economic analysis. We now show how syntax-oriented solutions (blocking connections, blacklisting certain known servers, filtering certain headers, etc.) do not address the full problem. Syntactic solutions rarely go beyond the first hop of communication whereas the multi-hop flow of data that is used over a longer period of time necessitates a semantics-based solution. The Do Not Track mechanism (which is expressed via a syntactic HTTP header specification) goes further as it tries to impose a longer term requirement on compliant servers but currently lacks a mechanism for checking compliance (other than threat of litigation). Recent press reports indicate that the effort has stalled. A semantics-based approach that addresses all aspects of the privacy conundrum is required. We illustrate the need for this through a couple of real examples.

The first example shows a complex sequence of events, the presence of multiple parties including some hidden ones, and old protocol decisions—all leading to leakage of private data. A users visits a popular Website (`www.AGEGROUPS.site`) which triggers fetching of `http://metrics.AGEGROUPS.site/...` (please see Section 3.1.1 of [12] for more details).

```
GET http://metrics.AGEGROUPS.site/...
Referer: http://www.AGEGROUPS.site/
Cookie: ...e=jdoe@email.com&f=John&l=Doe&...
```

The new URL appears related to the primary website visited by the user (based on the second-level domain name—`metrics.AGEGROUPS.site`). However, if we examine the authoritative DNS server of `http://metrics.AGEGROUPS.site`, we see that it actually belongs to a popular aggregator site.

Next, thanks to the way cookies operate in HTTP, the cookie associated with `www.AGEGROUPS.site` is sent to *any* site if the second-level domain name in the URL matches (as it does here). Further, the cookie actually placed by the primary website `www.AGEGROUPS.site`, as the *Cookie* header shows, includes personal information (name, email address).

Such a possibly unintentional sharing by the first party site is unknown to the user. This is further compounded by the fact that the same (hidden) aggregator may have access to the user's email address on her OSN account, thus enabling trivial linkage. It is also possible that the aggregator might not link this information with information received from other sites. The disclosure and the potential for linkage is invisible to most users as are the policies regarding linkage of personal information.

This scenario shows the need to identify all the parties involved, the data that is carried through automatically (first-party cookie) and understand the HTTP protocol's embedded decision to send first party cookies to any site whose URL matches the second-level domain string. The simple privacy protection scheme that blocks third party cookies (say) would not solve this problem due to the hidden semantic complexity of how cookies behave and how data bleeds across seemingly related sites.

Currently, OSNs groups a large number of privacy settings, assign default values, and decide the duration of their validity. Studies have shown that the complexity lead most users to use the default. Although users can change the values, OSNs regularly introduce new features affecting user's privacy allowing users only to opt out. When it comes to external entities (aggregators, applications), the information and choice provided to the users is limited. The communication between external entities and OSN is opaque to the users who cannot comprehend the flow of their data and remain unaware of actual and potential leakages.

The second example shows the sharing of unnecessary data, absence of control over it, the role of past privacy settings, and the limited responsibility undertaken by the OSNs. Consider a user installing an external application on her OSN account who is prompted for access to a set of permission bits. The user could accept the request or refuse and be unable to use the application. While a few applications ask for specific permissions at the actual time of usage, most ask for permissions at installation time without any accompanying rationale. Most users blindly accept the request for access. Separately, the user early on may have allowed for some permission bits to be globally accessible by all applications. This raises the possibility that the user's "consent" at installation time is not fully informed. An alternative idea of prompting users periodically may force some users to pay attention but annoy many users; this underlines the tussle between usability and improving privacy.

Scenarios such as sharing across OSNs based on old permission settings are even more complex. None of the protection mechanisms (including DNT) apply to information the user has *voluntarily* supplied to an OSN. Nor do they deal with the offline linkage which occurs mainly because of the prevalence of globally unique identifiers (which may be simple strings or a more complex bit vector consisting of various pieces of a user's personal information). Locating the minimal combination of bits that can lead to re-identifying a user is difficult. Syntactic solutions for prevention of leakage are incapable of predicting, let alone preventing even such routine occurrences. What is needed is a holistic examination of the full semantics of information sharing: flow analysis that examines past actions, current interactions, and possible future linkages.

6. TOWARDS A SEMANTICS-BASED APPROACH

Ideally, users would indicate at a high level the desired ex-

[9] http://www.oecd.org/dataoecd/8/51/46968784.pdf

[10] `http://www.sciencedirect.com/science/article/pii/S0165176512002182`

tent of privacy protection, which would then be translated into the appropriate underlying syntax in the OSN. In practice, this requires users to be aware of the leakage scenarios. Additionally, conveying intentions and having them interpreted unambiguously is hard. A more practical approach would allow users to visually see who has access to their actions, for how long, and possibly even control the context in which it could be used. To narrow the gap between syntax and semantics the interaction between the parties must thus be made significantly more transparent. The users must actively participate in the meta-dialogue and learn what happens to their privacy as a result of their actions. It should be possible to ensure that all the personal bits shared are actually required and to then trace their flow through the external entities. The OSN could help in increasing transparency and engagement of users so they can attain their individual desired level of privacy.

Our focus is on practical to narrow the gap between syntax and semantics in popular OSNs. Towards this end, we discuss preliminary tools. They are presented to illustrate the minimal steps necessary to move beyond static syntactic settings toward a dynamic analysis of the range of data spread and leakage possibilities.

6.1 Bridging perception and reality gap

Recently, we have developed a simple Facebook extension called Privacy IQ which poses a series of questions about the privacy settings associated with a user's objects (photos, postings etc.). Example questions include asking who all can see the list of a user's friends or who can see a photo posted on their page ("You, Friends, Friends-of-Friends, Everyone?"). Another asks if the user can "tag" someone in a photo even if they are not friends with them. If the user provides an incorrect answer, the app suggests how they can change their privacy settings to match their perception thus narrowing the gap between perception and reality. A privacy score is presented at the end.

Users can see their past privacy settings and the connections resulting from their settings. They can also easily see how their account is seen by others. Until Facebook's recently introduced Timeline feature made this easier it was quite cumbersome to see content from the past. A reflexive view enables a different perspective to users who can better comprehend access to their objects, learn the reach of their social graph, and the privacy implications of the applications they have installed. OSNs could encourage all their users to take similar surveys periodically. Additionally, simply being queried about the basic privacy model allowed users to learn more about it.

Privacy IQ aims to visually display the privacy reach of user's objects and the effect of their past privacy settings. By showing the set of entities who are able to access the results of their past and future actions, users are able to control the impact of their actions. Although the settings that they can change are limited to what is available in the OSN today, the users have a better understanding of the accuracy of their privacy perception and the need to periodically check the results of their past actions.

Privacy IQ is structured as an interactive survey to increase the potential for more users to take it. After the survey the user is asked for permission to post their score on their wall and share it with their friends. The aim is to take advantage of the potential for competitiveness amongst friends and induce them to take the survey and get a better score. The incentives to take such a survey includes desire to improve one's privacy, compete with their friends, and warn other friends about the lessons learned from their most egregious misunderstandings.

A couple of hundred users have taken the survey thus far and have reported gains. The common refrain has been pictures posted in the past with broader permissions than warranted, absence of clarity on the reach of their data (meant to be accessible just to friends were often mistakenly available to a much broader audience), and the fact that more fine-grained privacy controls were available but not used (individual pictures in an album could have more restrictive permissions). It should be stressed that Privacy IQ work is preliminary and work on obtaining a significantly large user sample is ongoing.

Privacy IQ addresses the longitudinal issue: users forget their original settings and do not realize that past permissiveness can impact future objects shared on the OSN. By forcing users to examine their settings we focus their attention on the reach of their permissions and spread of their data.

6.2 Addressing external linkage

Startups dealing with user reputation offer to correct offending information found by scraping visible information about users. OSNs can help by ensuring that no additional bits are contributed to external aggregators enabling linkage. OSNs can proactively ensure that only the absolutely essential data about the user is made available instead of shifting this burden to the user. A comparative analysis of applications could be done to order external applications from a privacy perspective.

OSNs could also increase the anonymity of user data by passing on only generic demographic information that still benefits advertisers. Obfuscating the data would improve privacy while still allowing targeted advertisements that are based on demographic information of users.

External linkage forces us to examine how parts of shared data may be combined with external information. Such combinations are not within the scope of syntactic protections.

7. CONCLUSION

As more users on the Internet share more information with each other as well as with commercial entities and data aggregators, the battle lines appear to be drawn between privacy advocates and consumer protection agencies on one side and the advertising industry and OSNs on the other side. The tussle to enable responsible sharing with reasonable privacy guarantees has devolved into a cat and mouse game. Research has yet to come to grips with the growing manners of data sharing through multiple devices. The complexity of providing highly usable privacy protection has also not been handled. The current syntactic protection methods do not capture the full reach and flow of user's data over time and across sites. Significant work remains in developing a more semantics-based approach. While the solution to the privacy conundrum includes serious legal and policy components, technology may also have much to offer.

Currently contributions by the privacy advocacy community are used by the few users who are proactively inclined towards privacy. Expanding the applicability of the tools to a larger audience is hard. Use of crowdsourcing to broaden

the reach is likely to be needed.

Meanwhile OSNs face negative publicity due to privacy leakage stories in the press and are spending millions of dollars in handling lawsuits and lobbying to avoid governmental mandates. Integrating some of the proposed techniques and encouraging users to better understand the reach of their data should help. Because OSNs have the most detailed knowledge about the interactions of their users, they are in a position to increase privacy requirements on external aggregators and applications. The increased role of OSNs and a semantics-based approach to privacy protection would let us know if colorless green ideas can indeed sleep furiously.

Acknowledgments

My thanks to colleagues for comments on a draft version of this paper: Vijay Erramilli, Sharon Goldberg, Ramesh Govindan, Saikat Guha, Maritza Johnson, Alan Mislove, Jeff Mogul, Jennifer Rexford, Rick Schlichting, Craig Wills, and Ben Zhao. My sincere thanks to Svetlana Yarosh and two anonymous reviewers whose comments contributed significantly in improving the paper. My thanks to Susan Landau for patiently shepherding this submission.

8. REFERENCES

[1] Marco Balduzzi1 et al. Abusing social networks for automated user profiling. In *RAID*, 2010. http://www.iseclab.org/papers/raid2010.pdf.

[2] Joseph Bonneau and Sören Preibusch. The Privacy Jungle: On the Market for Privacy in Social Networks. *WIES*, 2009. http://www.cl.cam.ac.uk/~jcb82/doc/privacy_jungle_bonneau_preibusch.pdf.

[3] L.A. Cutillo, R. Molva, and T. Strufe. Safebook: A privacy-preserving online social network leveraging on real-life trust. *Communications Magazine, IEEE*, 47(12):94 –101, 2009.

[4] Eszter Hargittai Danah Boyd. Facebook privacy settings: Who cares?, August 2010. First Monday, Volume 15, Number 8,http://www.uic.edu/htbin/cgiwrap/bin/ojs/index.php/fm/article/view/3086/2589.

[5] Datta et al. Understanding and protecting privacy: Formal semantics and principled audit mechanisms. In *ICISS*, 2011. http://www.andrew.cmu.edu/user/danupam/datta-iciss2011.pdf.

[6] De Cristofaro et al. Hummingbird: Privacy at the time of Twitter. In *Proceedings of IEEE Symposium on Security and Privacy*, May 2012.

[7] Ardagna et al. Exploiting cryptography for privacy-enhanced access control: A result of the prime project. *Journal of Computer Security*, 18(1):123–160, 2010. http://spdp.dti.unimi.it/papers/JCS2010-PRIME.pdf.

[8] Enck et al. Taintdroid: An information-flow tracking system for realtime privacy monitoring on smartphones. In *OSDI*, October 2010. http://static.usenix.org/event/osdi10/tech/full_papers/Enck.pdf.

[9] Felt et al. Android permissions: User attention, comprehension, and behavior. In *SOUPS*, 2012.

[10] Mislove et al. You are who you know: Inferring user profiles in online social networks. In *International Conference of Web Search and Data Mining (WSDM)*, February 2010.

[11] Perito et al. How unique and traceable are usernames? In *Privacy Enhancement Technologies Symposiurm (PETS)*, 2011.

[12] B. Krishnamurthy, K. Naryshkin, and C. Wills. Privacy leakage vs. protection measures: the growing disconnect. In *Web 2.0 Workshop on Security and Privacy*, May 2011. http://www.research.att.com/~bala/papers/w2sp11.pdf.

[13] Balachander Krishnamurthy and Craig Wills. Characterizing privacy in online social networks. In *Workshop on Online Social Networks*, August 2008. http://www.research.att.com/~bala/papers/posn.pdf.

[14] Balachander Krishnamurthy and Craig Wills. On the leakage of personally identifiable information via online social networks. In *Workshop on Online Social Networks*, August 2009. http://www.research.att.com/~bala/papers/wosn09.pdf.

[15] Balachander Krishnamurthy and Craig E. Wills. Privacy leakage in mobile online social networks. In *Workshop on Online Social Networks*, June 2010. http://www.research.att.com/~bala/papers/pmob.pdf.

[16] Helen Nissenbaum. A contextual approach to privacy online. *Daedalus the Journal of the American Academy of Arts & Sciences*, 140(4):32–48, Fall 2011. http://www.amacad.org/publications/daedalus/11_fall_nissenbaum.pdf.

[17] U.S. Department of Health Education and Welfare. Records computers and the rights of citizens. Report of the Secretarys Advisory Committee on Automated Personal Data Systems. DHEW No. (OS) 73-94. Government Printing Office. July 1973.

[18] Sai Teja Peddinti, Avis Dsouza, and Nitesh Saxena. Cover locations: Availing location-based services without revealing the location. In *ACM Workshop on Privacy in the Electronic Society (WPES)*, October 2011.

[19] Amre Shakimov et al. Vis-ã-vis: Privacy-preserving online social networks via virtual individual servers. In *COMSNETS*, 2011.

[20] Emily Steel and Jessica E. Vascellaro. Facebook, MySpace Confront Privacy Loophole. http://online.wsj.com/article/SB10001424052748704513104575256701215465596.html, May 21, 2010.