Requirements Analysis for Privacy in Social Networks

Prachi Kumari

Fraunhofer Institute for Experimental Software Engineering
Fraunhofer Platz 1, 67663 Kaiserslautern, Germany
Prachi.Kumari@iese.fraunhofer.de

Abstract. The rise and growth of social networks can be seen as empowering the user to change website's content by posting information using minimal technical knowledge. But, this empowerment has resulted in loads of sensitive data being let out unprotected in the public domain. To ensure user privacy, we need to understand privacy requirements relevant to social networks, per se. In this paper, we address this problem. We identify all the major stakeholders and their assets. We also look into the various aspects of user data that these stakeholders can be interested in. We then show how interests of various stakeholders can conflict and become threats for them. To counter these threats, we present a set of system requirements mapped to the respective privacy requirements.

Keywords: Data, privacy, requirements analysis, social networks.

1 Introduction

The concept of social networking is based on sharing information with people who have similar interests, hobbies, values etc. As a result, a lot of personal data[1] is conscientiously let out in the public domain by the users. This data can be of interest to various entities inside and outside of the social network. This exposes users to various kinds of threats.

Data protection in social networks is different from that in any other information system like in health or financial services because, firstly, users are ignorant of how data flows in the network, who all can access that data, and the different ways in which this data can be used (or misused) by others. Yet, they deliberately give out as much information as possible for adding Friends or becoming popular[1,2]. Secondly, there are many stakeholders who want access to user data and it is not difficult for them to do so (unlike in a bank, for example, where data access is strictly restricted). Lastly, various stakeholders have conflicting interests in data and it is important to ensure a balance of their interests. They cannot be ignored because they provide data, revenue, infrastructure and other support, and are therefore important.

[1] In this paper, Data refers to user data and Privacy refers to user privacy.

The existing works in this area present privacy requirements for users. They talk about the large amount of personal data that can be misused. But they do not provide any insight into the various aspects of data or the stakeholders who might be interested in these aspects of data. This restricts the scope of requirement analysis. Also, the existing studies do not bridge the gap between privacy requirements and system requirements.

In this paper, we work towards closing this gap. We present a more fine-grained analysis of privacy requirements. The paper proceeds in the following way: We identify the various stakeholders and their assets. Our argument is that data needs to be protected for the users' privacy and for the benefit of other stakeholders. As our motive is to achieve data protection, we analyze user data and present its various aspects which might be of interest to the stakeholders. Considering privacy requirements by including all the perspectives of data helps us to come up with system requirements for those privacy issues which could not be addressed till now, for example, unwanted disclosure by other users. We have identified all the threats to the stakeholders and their assets, based on the major studies and directives in this area. To counter these threats, we present a set of system requirements which we map to the privacy requirements. We also suggest organizational measures that can be enforced by the social networking service providers to counter insider threats. The paper concludes by a synopsis of the limitations of data protection in online social networks.

2 Related work

A lot of work in this area concerns the users' perception of privacy in online social networks. According to Acquisti and Gross, users reveal a lot of information on their websites even when they are not very knowledgeable about the security features[1]. Similar concerns are expressed by Edwards and Brown while discussing the threats of default settings in social networking websites[2]. Krasnova et al. looked into the motivations for this kind of user behaviour and are of the view that users disclose information about them for initiating and maintaining relationships and they trust the platform providers.

While Dwyer and Hiltz say that users trust the site and many of them also extend the online relationships beyond social networking[3], Krasnova et al. have shown that in wake of social and organizational threats, users withdraw from disclosing information about themselves or can resort to providing false information.

Williams et al. are of the view that users' behaviour in the social networks exposes them to various risks. Their study shows that users of different age group behave differently in terms of disclosing personal information online[4]. Stephan Weiss says that threats to privacy differ widely among individuals and are context-specific[5].

All these studies present privacy requirements for the users. However, opinions are divided among how privacy can be ensured. Edwards and Brown believe that technology and law together can provide data protection[2] whereas, James

Grimmelmann says that technical control won't work in the case of social networks[6].

Social networking has also gained the attention of administrative and legal organizations. The Trans-Atlantic Consumer Dialogue, European Network and Information Security Agency (ENISA) and the Data Protection Working Party of the European Commission recently came up with recommendations and opinions in this matter[7-9]. These directives cover various legal and social aspects of privacy and consider the social networking service (SNS) provider responsible for protecting the rights of users in online social networks.

These studies are different from our work because we present the privacy requirements from more than one perspectives. While our work builds upon the threats identified by them, we extend the privacy requirements to express them vis-à-vis system requirements. Also, the thorough classification of all types and aspects of data in social networks is an important contribution. To the best of our knowledge, we did not find any existing work that takes a data-centric approach on this issue. An overview of our work in this paper is shown in Figure 1 where the portions in gray are the focus of existing works.

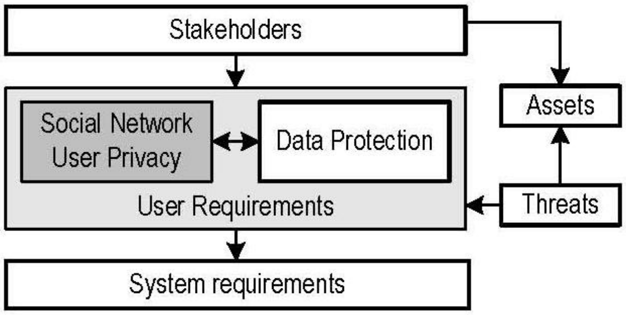

Fig. 1. Privacy requirements in social networks

3 Stakeholders in Online Social Network

We have identified *all the major entities who (can) get access to user-data directly or indirectly,* as stakeholders. They are shown in Figure 2. The arrows and their directions show the flow of user data and services between the stakeholders. This diagram is for an overview and does not identify all the data flows. For example, in order to offer their services, third-party application developers also provide data about themselves when they register with the SNS operator. But, this data flow is not shown here.

3.1 User

All persons(s) or group(s) of persons who participate in social networking through the Internet come under the class of *Users*. In Figure 2, they have been depicted as one entity. Users feed many types of data such as personal details, photos, videos, etc. into the social network. These data are prone to unauthorized access and use within the network as well as outside of it.

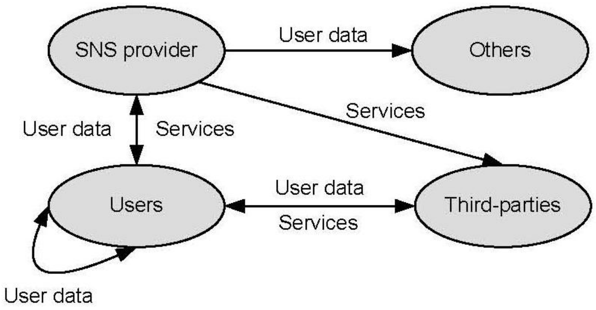

Fig. 2. Stakeholders in online social networks

3.2 Social Network Service (SNS) Provider

Also referred to as Social Network Operator (SNO), it is that entity which provides social networking service to the users. SNS providers generate much of their revenue through advertising which is served alongside the profiles and web content accessed by users. Users who post a lot of information about themselves on their profiles such as details of their families and friends, their profession, hobbies and other interests etc. offer a refined market to advertisers wishing to serve targeted advertisements based on that information. Another revenue-generation model used by some SNS providers is to offer two types of membership and charge a registration fee from users wishing to have a privileged account. For example, this model is used by XING[10].

3.3 Third Parties

Third parties connect to the users and SNS providers for purposes other than social networking. Their connection with the users can be classified as:

- Directly, as developers and providers of plug-in/add-on applications on the social networking website
- Indirectly, as advertisement agencies, headhunters etc.

While third-parties get access to valuable user-data, there is no proper mechanism to check unauthorized access and usage of data by them. A look at the Developers Terms and Conditions documents of the major SNS providers points to this [11-13].

3.4 Others

All those who are not directly involved in social networking as Users or providers come under this class. They have varied interests in user data such as monitoring, law enforcement, business or other objectives like research & development education etc. Some examples of *Others* can be law enforcement agencies, research organizations, employers, universities, governments and secret services.

In spite of not being active online, *Others* also face threats in the social network. They can be victims of unwanted disclosure or defamation. We talk about threats in Section 4.2. But before that, we look at the various assets which can be targeted by attackers and can face threats in social networks.

4 Assets and Threats

4.1 Assets

In this paper, we define asset as *a tangible or intangible resource of an entity, which has some value for that entity and, which can be exploited by anyone within or outside the social network.* Based on this definition, we have identified a list of assets with the corresponding stakeholders, shown in Table 1.

Assets	**Stakeholders**
User data	Users, SNS providers, Third-parties, Others
Devices	Users
Physical safety	Users, Others
Reputation	Users, SNS providers, Third-parties, Others
Revenue	SNS providers, Third-parties

Table 1. Stakeholders and their assets

As privacy is the right of an individual to exercise control over his data [14], users can lose their privacy when they participate in online social networking because they lose this control[2] over their data. Providing users more control over their data can help achieve privacy in social networks.

[2] Privacy is the claim of individuals, groups and institutions to determine for themselves, when, how and to what extent information about them is communicated to others.

Besides, data is an asset for all the stakeholders. Therefore, it is very important to understand the various types and aspects of data.

User Data: Figure 3 gives an overview of all kinds of user data in social networks. It is broadly classified into two:

- **Payload data**: All data posted by the user e.g., personal profile data, videos, images, messages etc.
- **Traffic data**: All data generated by users' activities in the social network e.g., user's IP address, browser and OS specifications, search terms or profiles searched, pages visited etc.

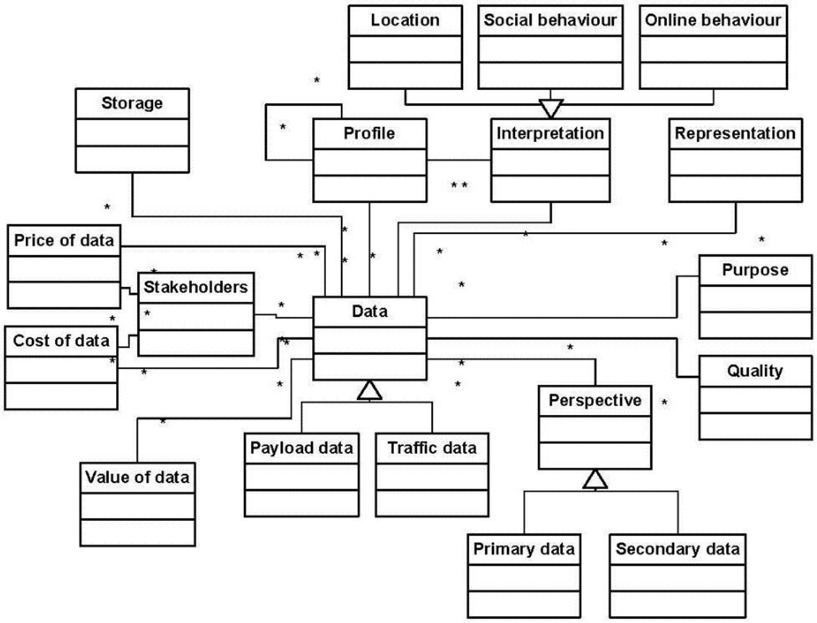

Fig. 3. Overview of data in a social network

We can further classify data as primary and secondary data, according to **perspectives**. Thus, we have four types of data in a social network:

- **Primary payload data**: All data posted about a user by himself. For example, the data posted by user while creating his profile on the website.
- **Secondary payload data**: All data posted about the user, by other users. For example, another user can post a picture of this user in his album and link it to him.
- **Primary traffic data**: All traffic data generated by the user while surfing. For example, the profiles visited by the user.

- **Secondary traffic data**: All traffic data generated when other users visit this user's profile, view his pictures, videos etc.

Figure 3 also illustrates the following different aspects of data in a social network.

- **Profile**: Can be one or many users' profiles. A single user's profile can be further classified as 'behavioral profile' and 'uninterrupted personal history'.
- **Interpretation**: For example, the user's traffic data can be interpreted to infer that he likes to make friends with 'single, independently-living persons'.
- **Representation**: Data can be text, images, audio or video files.
- **Storage**: Details of how data is stored, where it is stored and for how long it is stored, i.e., 'persistence' and 'location' of data.
- **Quality**: How correct and updated the data is. It can be further classified as 'freshness' and 'correctness' of data.
- **Cost, Price, Value of data**: *Cost* of data is required to collect/generate the data, *price* is that amount for which someone can buy this data and the *value* of this data depicts the financial potentials of the data if it is used, sold etc. All these factors are important for various stakeholders.
- **Purpose**: The purpose of data access or collection. In some jurisdictions like the European Union, purpose of data is very important for legal reasons.

Privacy requirements for data vary according to these aspects. For example, friends of a user might like to know what this user bought recently (freshness of data), but, an advertisement provider would be more interested in similar records collected over the past six months (behavioural profile or uninterpreted personal history). While the user might allow his shopping information to be visible to everyone, he might not want it to be stored as history. Therefore, he might not want to control data access but would like to control unauthorized storage and use of it.

Another important point in this context is that many stakeholders want control over the same data for various interests. For example, third-party advertisement providers want access to user-data so that they can serve tailored advertisements to the users. SNS providers encourage users to post as much details about themselves as possible so that they can make more money by selling these data to the interested third-parties. But the user wants privacy and does not want to give control over his data to others. In this case, the interests of stakeholders conflict and can pose a threat for the user. For example, posting too much information about the daily activities of the user can be good for the SNS provider but it can lead to the user being stalked.

There are also other assets of the stakeholders in social networks. These assets can also be affected by threats to user data. For example, a user's physical safety can be threatened or an SNS provider can lose reputation and/or revenue because of data compromises. Therefore, data protection is important for the safety of all the assets of all the stakeholders. However, in this paper we do not go into the detailed analysis of other assets. Instead, in the following sections we look at the threats to these assets and how these threats can be countered.

4.2 Threats

For this paper, we use the following definition of threat:

A potential for violation of security, which exists when there is a circumstance, capability, action, or event that could breach security and cause harm [15].

Based on the above definition, we have identified the prominent threats along with the affected assets and the stakeholders. As is evident from Table 2, threats to data also affect the physical safety and reputation of the users. This can result in lawsuits, revenue-loss or other threats for the other stakeholders. Thus threats to data affect all the stakeholders and their assets.

System requirements depend upon the source of threat and the aspect of data affected. For example, to counter identity theft, SNS provider needs to have some identification mechanism at the time of registering users. This can prevent an individual to create a fake profile on someone else's behalf. But, to prevent identity theft by/of the registered users, there needs to be a mechanism that checks for multiple profiles with same identification details. So, for one particular privacy requirement, we can have many system requirements.

Threats	**Assets**	**Stakeholders**
Digital Dossier Aggregation	Data, Physical safety	Users
Identity theft	Physical safety, Reputation	Users, Others
Behavioural profiling	Data, Physical safety, Reputation	Users
Unwanted disclosure	Data, Reputation	Users, Others
Defamation	Reputation	Users, Others
Stalking	Physical safety	Users
Cyberbullying	Physical safety	Users
Surveillance	Physical safety	Users
Malicious attacks	Devices	Users
Legal action/lawsuits	Reputation, Revenue	Users, SNS providers, Third-parties
Ban	Reputation, Revenue	Third-parties, SNS providers
Loss of revenue	Revenue	Third-parties, SNS providers

Table 2. Stakeholders, assets and threats

By now we have established that data is an important asset for all the stakeholders and it needs to be protected for countering the threats to users and other stakeholders. In the next section, we show how data can be protected by technical means, that is, we present system requirements for the identified privacy requirements.

5 System Requirements for Privacy

We argued in Section 4.1 that users should be given more control over their data to protect their privacy. This requirement has been further broken down as control over access, usage and disclosure of data. Each privacy requirement should be fulfilled by a set of system requirements. These have been identified below.

5.1 Access control requirements

User should be able to control the visibility of his profile and other data posted by him.

System requirements

- Only registered users should be able to access profiles and provide third-party application services to other users.
- Default settings in profiles should be revised to provide more opt-ins than opt-outs for sharing data.
- Privacy settings for profiles should have different levels of granularity to adapt to different user-preferences. The existing privacy settings in the social networks based on the notion of *friends* are insufficient. A better approach can be to include trust models for categorizing the contacts and map different categories of contacts and data to permissions.
- User should be able to know who sees what content about him.
- User should be able to select which profile information he wants to release for using a third-party application.
- Harvesting profiles using automated tools like scripts should not be allowed. Effective technical mechanisms should be used to stop malicious users or other hackers.
- User should have option to specify the interest for getting updates, news or advertisements instead of getting them by default. This can be a better model also for the advertisement providers as the advertisements would target those who are *actually* interested in them rather than those who *might be* interested.

5.2 Usage control requirements

User should be able to control unauthorized storage and use of his data.

System requirements

- Right click options should be disabled to control unauthorized actions.
- Images, videos and audio files should not be downloadable in any format.

- Making copies of displayed information (e.g., copy-paste, save, print functionalities and taking screenshots) should be disabled. One way to do this can be to introduce a policy-based mechanism of data usage control where the user can specify how he wants his data to be used (or not used) and then enforce these policies when operating the social network.
- Criteria for banning a user or third-party for policy violations should be stricter.
- Persons should not be allowed to create multiple profiles of them.
- Semantics of the operations on data should be clearly defined. For example, updating content can mean that the old content is overwritten by this new one instead of making a separate copy of new content and hiding the old copy. Similarly, deleting a content can mean that the content is deleted from all the places in the system, at the SNS provider's end also, and not merely made invisible to users.

5.3 Disclosure control requirements

User should be able to control that no one posts any content about him without explicit permission.

System requirements

- There should be effective verification mechanisms to prevent identity related frauds and fake profiles. One such measure can be user-verification by mobile number instead of the email. Although not fool proof, this would be more effective than the existing mechanism of verification by email.
- Registration mechanism should ensure that automated programs are not able to clone profiles or register as users.
- Tagging or linking a user should require permission from the user.
- Posting content about anyone should require permission from that person/group. User should know who posts what content about him.
- There should be a mechanism that can check profiles for duplication of identification data.
- If some user tries to post content about someone who is not registered with the social networking website, he should not be allowed to do so.

The aforementioned system requirements have been developed as delta of the existing system functionalities of five popular social networking websites: Facebook, MySpace, Orkut, Xing and StudiVZ. We are aware that these system requirements might not be comprehensive in general, but they are sufficient to counter the threats identified by us in Table 2. As these threats have been identified based on an extensive study of the scholarships in the area, they represent a comprehensive picture of threats and privacy requirements and thus the system requirements expressed here are comprehensive in the context of our argument in this paper.

Intuitively, technical measures can become more effective when supported by organizational standards and practices. In the next section, we suggest some organizational measures for privacy protection by social networking service providers.

6 Organizational Measures

Organizational measures are needed to meet legal requirements, to protect data from misuse by internal and external employees of the SNS providers, to inculcate trust in users by improving the transparency of data usage, storage and flow and to achieve competitive edge in the market. Following is a synopsis of the organizational measures which should be implemented by the SNS providers.

- *User policy*[3] should not be changed frequently and whenever there is a change, the user should be informed of the same.
- User policy should state the purpose of data collection clearly and in adherence to the governing laws.
- User policies should inform users which mechanisms are used to prevent and handle cases of data misuse by other users, third-parties and the service providers' employees.
- SNS providers should make the user policies simple enough to be understood by the average user.
- SNS's organizational policy should include measures for protecting user data from unauthorized access and usage by their employees.
- To assure the users of their privacy, SNS providers should include 'data and privacy protection' as a key value of their organization.

In spite of all the suggested measures to protect data in social networks, there are some practical limitations to achieve complete data protection in social networks. These limitations arise out of the human factors associated with any automated system. For example, function creep of data by trusted humans cannot be handled only by technical measures. Also, studies show that there is always a trade-of between the refinement of control and understandability of the system [16], i.e., if the user is given more fine-grained control, the system becomes more complicated for him to use. As privacy requirements give more control to the user over his data, human factors and the associated issues become very crucial to the success of any privacy mechanism. In the following section, we talk about two such kinds of limitations.

7 Limitations of data protection

The limitations in enforcing a data usage control policy can be posed by technological developments, law, user requirements, business interests and values.

7.1 Legal Limitations

- The spread of social networks is across boundaries whereas laws in different countries do not treat the relevant issues in similar fashion.

[3] The term User Policy is used to denote Terms and Conditions accepted by a user at the time of registration.

- Rights of various stakeholders in the social network differ according to jurisdictions.
- In some cases, different laws within a country are in conflict.
- If perpetrator and victim are in different countries, inconsistency in laws can be used by the perpetrators for their advantage.
- Cyber laws across various countries are in development phase and there is almost no clarity regarding 'ownership' of the data in many contexts.

7.2 Users and usability Related Limitations

- Implementing data protection might complicate the applications. This can make usability a major issue for data protection mechanisms. SNS providers (e.g. Facebook) have undergone criticism for such reasons.
- Users do not generally like rigid solutions. They might consider data related threats as predictions and not reality. Users view privacy in social context rather than in terms of access lists and file permissions. Therefore, they can be easily lured into overriding security measures[6].
- Privacy mechanism has to be implemented with certain assumption of the level of privacy it would provide to the users. If this level is too low, it would be equivalent to having no data protection by default and would rely on user to understand and set the privacy levels. Given the fact that a large number of users keep their privacy settings on by default, this won't help much. If the privacy settings are by default very high, the users would find it hard to socialize and would turn off the settings treating it as a bug.

7.3 Conclusion and Future Work

The aim of this paper is to present a comprehensive requirement analysis for privacy and bridge the gap between the user and the system requirements. We started by identifying the major stakeholders and their assets. We presented an analysis of data and argued that protecting data is important for all the stakeholders and their assets. Then we looked into the various threats that can endanger them. To counter these threats, we have presented a list of system requirements which have been mapped to the corresponding privacy requirements. The set of system requirements is relevant to the context of this paper and also applies to the general case.

As technical solutions alone are not sufficient to address this problem, we have also suggested some organizational measures. We have also analyzed the approach of protecting data in general and discussed the various limitations of data protection in social networks.

This work is the first step towards developing a data protection mechanism to ensure user privacy in social networks. The next step in this direction is to come up with a further refinement of requirements in terms of implementation details. Based on a detailed requirements specification, we would design, implement and test a data protection mechanism for a set of users. This can be done at various levels in the system. Enforcing privacy requirements at the applications layer can be a good start in this direction. As most of the web based social networks are accessible through web browsers, we plan to enforce some of the requirements mentioned in this paper for Mozilla Firefox browser. We also introduce a trust model to replace the existing relationship model of most of the popular social networks of today. This work is in progress and the results are awaited. Based on the evaluation of the enforced mechanism, we plan to extend the solution to other layers in the system.

References

1. Alessandro Acquisti and Ralph Gross. *Imagined Communities: Awareness, Information Sharing, and Privacy on the Facebook.* In Privacy Enhancing Technologies Workshop (PET), Robinson College, Cambridge, United Kingdom, June 2006.
2. Lilian Edwards and Ian Brown. *Data Control and Social Networking: Irreconcilable Ideas?* HARBORING DATA: INFORMATION SECURITY, LAW AND THE CORPORATION, A. Matwyshyn, ed., Stanford University Press, 2009.
3. Catherine Dwyer and Starr Roxanne Hiltz. *Trust and privacy concern within social networking sites: A comparison of Facebook and MySpace.* In Proceedings of the Thirteenth Americas Conference on Information Systems, Keystone, Colorado, USA, August 2007.
4. Kaven Williams, Andrew Boyd, Scott Densten, Ron Chin, Diana Diamond, and Chris Morgenthaler. *Social Networking Privacy Behaviors and Risks.* Seidenberg School of CSIS, Pace University, USA, 2009.
5. Stefan Weiss. *The Need for a Paradigm Shift in Addressing Privacy Risks in Social Networking Applications.* The Future of Identity in the Information Society, 17th June 2008.
6. James Grimmelmann. *Facebook and the Social Dynamics of Privacy.* Legal Studies, New York Law School, 7:33{34, 2008/2009.
7. ARTICLE 29 - Data Protection Working Party, Adopted on 12 June 2009.
8. Giles Hogben. *Security Issues and Recommendations for Online Social Networks.* Technical report, ENISA, October 2007.
9. *Trans Atlantic Consumer Dialogue: Resolution on Social networking.* http://www.tacd.org/index2.php?option=com_docman&task=doc_view&gid=208&Itemid=40, May 2009.
10. Xing homepage. http://www.xing.com.
11. Facebook Platform Guidelines. http://wiki.developers.facebook.com/index.php/Platform_Guidelines, 26 September 2009.
12. MySpace Apps Developer Addendum to MySpace.com Terms of Use Agreement. http://wiki.developer.myspace.com/index.php?title=Developer Addendum to MySpace.com Terms of Use Agreement,26 September 2009.
13. Orkut Terms of Use of Applications. http://www.orkut.com/AppTerms.aspx, 26 September 2009.
14. Alan F Westin. *Privacy and freedom.* Atheneum; 1st edition , 1967.
15. William Stallings. *Cryptography and network security: principles and practice.* Prentice Hall, 4 edition, November 26 2005.
16. M. Wegdam and D-J. Plas. *Empowering users to control their privacy in context-aware system through interactive consent.* Technical Report TR-CTIT-08-66, Enschede, December 2008.