Social Networking, 2013, 2, 157-164
http://dx.doi.org/10.4236/sn.2013.24015 Published Online October 2013 (http://www.scirp.org/journal/sn)

Taxonomy for Privacy Policies of Social Networks Sites

Sergio Donizetti Zorzo, Rodrigo Pereira Botelho, Paulo Muniz de Ávila
Computer Science Department, Federal University of Sao Carlos, Sao Carlos, Brazil
Email: zorzo@dc.ufscar.br, rodrigo_botelho@dc.ufscar.br, paulo.avila@dc.ufscar.br

Received July 2, 2013; revised August 7, 2013; accepted September 25, 2013

ABSTRACT

Social networking sites (SNSs) are becoming increasingly popular on the Web. Sharing personal information in these networks can be dangerous, considering that malicious users can get access to this information and use them for purposes other than the original. Although SNSs typically provide tools for users to set who can access their shared data, this access restriction only applies to network users and not for third parties and the social network itself. In this paper, we present both a mechanism to enhance privacy in SNSs and taxonomy for classifying SNSs privacy policies. We combine and extend two taxonomies of privacy, unifying them to classify SNSs privacy policies and also the aforementioned mechanism. We evaluated the taxonomy classifying six SNSs privacy policies and the mechanism, presenting the results and our findings about the classification.

Keywords: Security; Privacy Policies; Content Protection; Privacy Taxonomy

1. Introduction

The social networking sites are becoming increasingly popular, and the best known, such as Facebook, Twitter and Orkut, have dozens or even hundreds of millions of users. The adoption of social networking sites seems to be a trend as other means to access such services becomes available.

Given this large amount of user information, the social networking sites protect users data through its privacy policies and security measures and privacy tools, trying to get users confidence in using the service. The privacy tools is one of the features that contribute to maintaining the privacy of users is the privacy settings provided by the sites, which implements the sites privacy policy. This is usually a section of the site where the user can set the visibility of data that will be provided to the network. But even with this feature, users of social networking sites are still revealing their private data in a dangerous manner [1].

The target of this paper is present a mechanism to enhance privacy in SNSs—protecting user privacy against web crawling and server invasion—and after this, we provide taxonomy for classifying the privacy policies of SNSs and the mechanism. For this, we combined and extended two previously proposed taxonomies of privacy. A taxonomy for social networks data [2] and a taxonomy of privacy in the juridical area [3]. The classification of privacy policies in the taxonomy helps us to better understand how they work, in addition to contribute to a formal and standardized comparison means. The usefulness of the taxonomy goes beyond of the classification of privacy policies and may be used to classify privacy related issues, as we show by classifying the privacy mechanism.

This paper is organized as follows: Section 2 presents related work, Section 3 presents the privacy mechanism, Section 4 presents the taxonomy, Section 5 describes the classification of the privacy policies of six social networking and the mechanism and finally, Section 6 presents our findings.

2. Related Work

The social networking sites are for the most part, similar with respect to the data that can be collected. In general, they are characterized by allowing the publication of a public profile, allowing other members in the network to identify the user in order to establish a relationship of friendship and enable navigation by members of the network [4]. It can be said that in order to characterize a social network, it is essentially necessary to have data and profile Data about the relationship between users. Other data are also used in social networks [3,5], as discussed in Section 3.

The SNSs privacy policies make references to these types of data to establish rules about what and how data will be collected and used. One way to evaluate and compare the privacy policies of different sites is classifying them into taxonomy. Some aspects are related, as the taxonomy of social networking sites data [3], about privacy [6,7] and the privacy policies themselves [5]. About the privacy enhancing tools, there are a number of researches that face privacy concerns in SNSs in context of social applications, access control and encryption.

Next, we present a privacy mechanism that protects user privacy and keep possible to provide personalized services, maintaining the Integrity of the Specifications.

3. Privacy Mechanism

The risk involved in sharing data is the maintenance them, because once the data were shared in the social network, may occur user lose control over their data. For example, a copy of the data can be made for third party services.

The privacy mechanism presented in this paper, called Privacify, makes use of encryption methods to encrypt data before it is sent to the social network. Thus, even if the data is copied to third-party services will not be very usefulness, since a recipient must know the keys to be able to read a message content.

Some requirements are essential to make the mechanism utilization in practice to interfere as little as possible on how a user uses the network. One of these requirements is the user ability to use this proposed solution from any computer. Another important requirement is to not trust the social network server even if it can be trusted. Thus, the data is shared only by those who were initially assigned by the data owner. Finally, a user should be free to return the normal use, without adding protection to data.

Figure 1 shows the components that make up the mechanism. The *Social Network* component is the representation of any social network which the mechanism will be applied. The *Browser* component is the representation of tool or program which user access and use the social network. This component is responsible for preparing the messages in the **Privacify-Message** format before it is sent to the social network. Finally, the *Keys Repository* component is a representation of a service that maintains information of user keys and the keys of the user’s friends. These keys are used to encrypt the message and then to read the encrypted message.

The components *Browser* and the *Keys Repository* are trusted elements in the mechanism and the arrows in **Figure 1** indicate the directions in which communication can occur. Note that the *Social Network* component does not communicate directly with the *Keys Repository*.

In a simplified manner, to a user be able to communicate using the *Privacify*, the user must generate a pair of public/private key and obtain the public keys of all users to whom he wants to maintain communication. These keys are stored in the *Keys Repository* and every time the user is using the *Social Network* these data are loaded into the *Browser*. It is the role of the *Browser* component to ensure that those keys are provided with security and privacy for the *Keys Repository*.

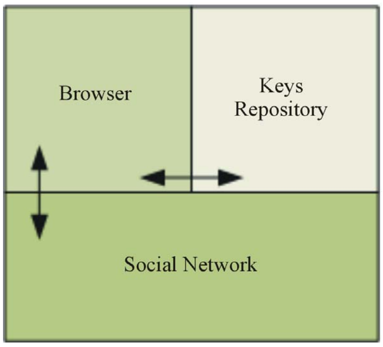

Figure 1. Mechanism components.

Every message sent to the Social Network server is on **Privacify-Message** format, which is illustrated in **Figure 2**.

The message can be divided into three sections: *Header*, *Encrypted Message* and *Aggregated Data*. The *Header* section contains information for that authorized users can read the message. In other words, it must contain sufficient information to each related user can be able to decrypt the cipher text. The *Encrypted Message* contains the payload of the message. To support advertising and access to specific data through social applications, the *Aggregated Data* field was added. With this field it is possible to aggregate some sensitive information, so the exact values are not revealed. For example, instead of providing precisely the age, we can put an age range in the *Aggregated Data* field.

The mechanism supports both messages sent to a single user or for multiple users. The difference between the two types of messages is the number of users listed in the message *Header*. The cipher text is unique not being necessary to encrypt the message *N* times to send to *N* users, which would make the mechanism implementation prohibitive for reasons of overhead in message size.

It is important to note that the proposed model does not guarantee the total privacy of user data. Social connection data, such as friends list, are still visible to the social network. However, ensures additional privacy through encryption of data that are posted explicitly. This additional privacy protects data from social networking and other sources, if any leaks.

Figure 3 shows the levels of privacy that can be obtained with *Privacify*.

The lowest layer is the level of privacy provided by the actual social network. It may include access control in parts of user data but this control is in relation to other users and not for third parties. The layers *Low*, *Medium*, *High* and *Custom* relate to levels of privacy provided by

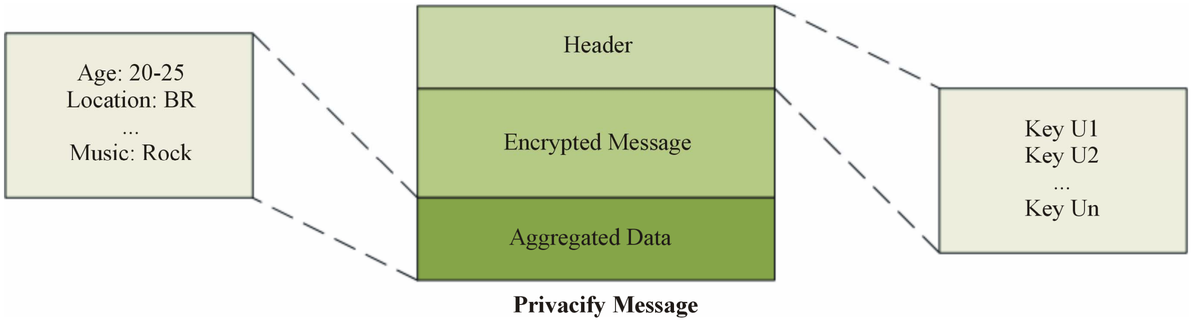

Figure 2. Message format.

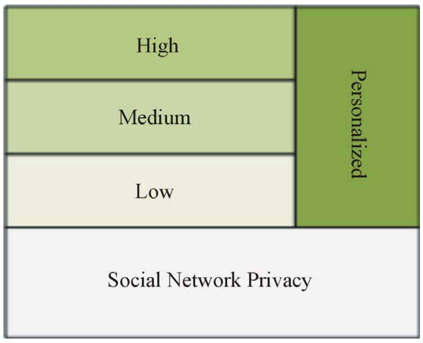

Figure 3. Privacy levels.

Privacify. With the exception of *Custom*, the top layers always provide protection from lower layers.

The *Low* layer of privacy only protects the user profile data, e.g. name, email, age, political views, among others. Some of these data, however, can be provided in the form of aggregated data for the social network in this way the services already offered are not harmed. *Medium* layer protects all text-based messages, for example, comments on photos, testimonials, among others. The protection provided by the *High* layer goes beyond text messages, providing privacy for all data posted by the user, for example, photos, videos and more. Finally, the layer *Custom* user can choose which data you want to keep private.

By observing the levels of privacy in **Figure 3**, one can see that the *Privacify* can be used to extend the privacy of online social networks supporting data privacy as well for third parties including the social network itself. To illustrate, suppose that a social network that allows users to change the policy on access to their profile data to "public". If user set up privacy level as *Low* on *Privacify* his profile data may also be retrieved by all users of the network, but only authorized users will be able to read the content so we are extending the privacy of social network.

We implemented the mechanism as a Google Chrome Browser extension for Orkut social network. The implementation details are discussed in the presentation of the Privacify [8].

4. The Taxonomy

This section presents two taxonomies used as dimensions of the new taxonomy proposed in this paper. In Subsection A, is presented the social networking sites data taxonomy proposed by Schneier [3]. In Subsection B, we present the taxonomy of privacy proposed by Solove [6]. This section is closed under Subsection C, presenting a new taxonomy that combines and extends the previous ones.

4.1. Social Network Data Taxonomy

Schneier [3] presents a list of six categories in which data available on the social network can be classified: Service data, Disclosed data, Entrusted data, Incidental data, Behavioral data and Derived data.

Service data are user-supplied data before it can access the service. These data are known as identifiable data, because they uniquely identify users on the system. Disclosed data are data that the user posts in his own page. These data are also known to form the user profile. Entrusted data are the data that the user posts the page to other network members. It is similar to Disclosed data, but the difference is that in some cases, after posting the content the user has no control over the data. Incidental data are data that other network members post about you. It is also similar to Disclosed data, but the difference is that it was not you who originally created the data and in some cases you have no control over them. Behavioral data are data that the site collects about the user's activities during its use. Derived data are derived data from the data aforementioned. The derived data can be generated using various techniques, such as data mining.

Comparing the proposed data above against the Wu *et al*. SNSs data [9], we find some similarity as illustrated in **Table 1**.This separation is important and helps to explain Information Collection dimension of on taxonomy, in Section 4.

The Registration data can be directly mapped for Service data, and Activity data can be mapped to Behavioral. However, Networking and Content do not have a well-defined mapping, only can be said that these data are

Table 1. SNSs data comparison.

Wu [9]	Schneier [6]
Registration	Service
	Disclosed
Networking Content	Entrusted
	Incidental
Activity	Behavioral
	Derived

Table 2. Data generation.

Data Generation	
User	Service
	Disclosed
	Entrusted
	Incidental
System	Behavioral
	Derived

relate to Disclosed, Entrusted and Incidental data. Additionally, Wu *et al.* claim that Activity data are derived and provided to third parties, but does not explicitly include the data type as is done with the Derived data type.

Another interesting question is related to data generation. One can clearly see that piece of data is generated by the user and part is generated by the system, as can be seen in **Table 2**. Data Service, Disclosed, Entrusted Incidental and are explicitly created by users.

Although Behavioral data are user's data about his activities on the system, the user does not provide in an explicit way, we classify it as generated by the system. Finally, the data are certainly Derived also generated by the system.

4.2. Taxonomy of Privacy

Solove [7] argues that is difficult to define privacy in a coherent and precise manner, showing that the privacy definitions often end up embracing aspects that are not only privacy and then end up not addressing all the aspects that are related to privacy. He introduces the taxonomy of privacy which is divided in four parts: Information Collection, Information Processing, Information Dissemination and Invasion.

The Information Collection has surveillance and interrogation. Surveillance is a way to get the data through observation, listening to or recording the user's activities. Interrogation is a way to get user data explicitly, for example, with the use of forms.

The Information Processing contains Aggregation, Identification, Insecurity, Secundary Use and Exclusion. Aggregation is the process of combining multiple user information in order to make the most complete data set. Identification consists in associate a data set with a specific individual. Insecurity is related to the failure to implement adequate security measures to ensure protection of user information against unauthorized access. Secondary Use consists in the use of user data with different purpose than originally intended. Exclusion is related to the lack of mechanisms to notify the user about who owns his data and to let him participate in its handling and use.

The Information Dissemination contain Breach of Confidentiality, Disclosure, Exposure, Increased Accessibility, Blackmail, Appropriation and Distortion. Breach of Confidentiality is to make public or share private information of the user. Disclosure is to reveal true information about a user which may change the way other members of the network judge him. Exposure involves the exposure of issues related to nudity, grief, among others. Increased Accessibility is to increase the possibility of access to user data. Blackmail is related to the threat to disclose your personal information. Appropriation is to use the user's identity to serve the interests of others. Distortion involves the dissemination of false information about the user. Invasions have intrusion and decisional interference. Intrusion is related to issues that disturb the solitude and tranquility of the user. Decisional Interference involves government intervention in decisions of the user regarding his private affairs. Although the above taxonomy was originally designed for the legal area, its use in other areas is not restricted, as shown below in subsection 4.3.

4.3. Our Taxonomy

The purpose of this taxonomy is to enable the classification of data available on social networking sites in terms of privacy. Therefore, the taxonomy can be used to, but not limited to, the classification of privacy policies and privacy mechanisms. For this, we combine the taxonomy of data social networking sites [6] with a taxonomy of privacy [9], placing each one in one dimension. Taxonomy organized into four sections, according to the taxonomy of privacy, information gathering, information processing, information dissemination and invasion.

The first part, Information Collection, is a bit different from the others; it presents in one axis the social networking sites and in the other axis shows social networking sites data, as shown in **Figure 4**. This part of the taxonomy aims to indicate how the data are obtained and can be done implicitly (Surveillance) or explicitly (Interrogation). Each pair {site, data} in this part has an associated value S or I or both, indicating how the collection is made for a particular data type in a specific site. We called Orkut*, the Orkut with the privacy mechanism presented in Section 3.

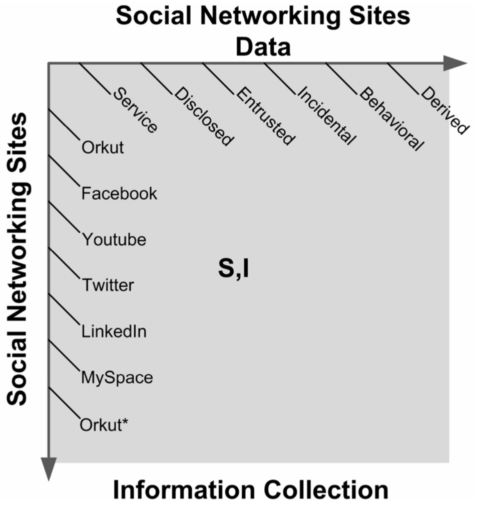

Figure 4. Part one: information collection.

The remaining parties have the taxonomy of privacy on one axis and data taxonomy onto another. Thus, instead of listing all the social networking sites in only one table, like in Information Collection, it's needed a table to classify each site, as illustrated by **Figures 5**-**7**.

Differently from what happens in Information Collection, each pair {information processing, data} has a given value associated with and may be House (H), Third Party (TP), Data Provider (DP), Users (U) or Not Allowed (N). More than one value at a given par is allowed. House is the social networking site, Third Party are third parties who may be advertisers on the site, developers of social applications for the site, among others. Data provider is the user who is providing the data and Users are other users of the network. By assigning one of the values cited above for the specific pair, we are actually classifying who is responsible for the privacy action for an item of data, if applicable. For example, if a social networking site has value H for the pair {Aggregation, Disclosed} this means that the site can aggregate user Disclosed data and if this happen does not means privacy invasion because the user is aware of such action. **Figures 6** and **7** illustrate, respectively, Information Dissemination and Invasion parts and are similar to the interpretation of Information Processing.

5. Privacy Policies Classification

This section shows the classification of the social networking sites privacy policies in the proposed taxonomy. As the classifications in Information Processing, Information Dissemination and Invasions need a table per SNSs, we decided to show here only classification of Orkut and Orkut* due space limitation.

5.1. Information Collection

The classification of the privacy policy of the sites in Information Collection aims to identify the way that data are collected. The taxonomy predicts data may be collected in explicitly (I) or implicitly (S).

According to **Table 3**, all sites use explicit collecting for user data—Service, Disclosed, Entrusted and Incidental data. For the automatically generated data implicitly collection is used. While all user data collection has been categorized as explicit, nothing prevents the use of implicit collection. For example, in the case of YouTube we can use a Google account to register for the service and thus at the registration data service could be obtained by an active user session if it were logged to your account Google.

5.2. Information Processing

The Information Processing part aims to identify some kind of modification or use made on user's data. When a site states in its privacy policy that makes some kind of Information Processing in the data, means that in doing such action would not be breaking the user's privacy, because in this case there is consent.

Unlike the classification of Information Collection, the Information Processing organizes the classification of each site in different table. Another point to note is that the data Derived, Disclosed, Entrusted and Incidental were grouped into a larger category called User Data. The same goes for the Behavioral and Derived data, which were grouped into a category Auto Generated Data. This categorization was needed, because it makes no sense to classify each data type individually, but for which the source data has been generated.

Below we discuss the classification of the site Orkut and Orkut*, for comparison purpose.

Information Processing classification of Orkut and Orkut* are identical, as can be seen in **Tables 4** and **5**. Orkut make explicit on their own privacy policies that can aggregate data from users for improved service, and to share this aggregate data with strategic partners for advertising purposes. However, this share does guarantee that any data are associated with a specific user, not allowing its identification. This way, the privacy mechanism enforces this policy.

Moreover, it is guaranteed by the privacy policy that data will never be used for different purpose other than the original, except in cases that represent a security threat to the network, other users or violates any law. Finally, Orkut ensure that the user can participate in the handling of your data.

5.3. Information Dissemination

Information Dissemination tells how the user data are shared on the network and how access to this data can be obtained. Just as the classification of Information Proc-

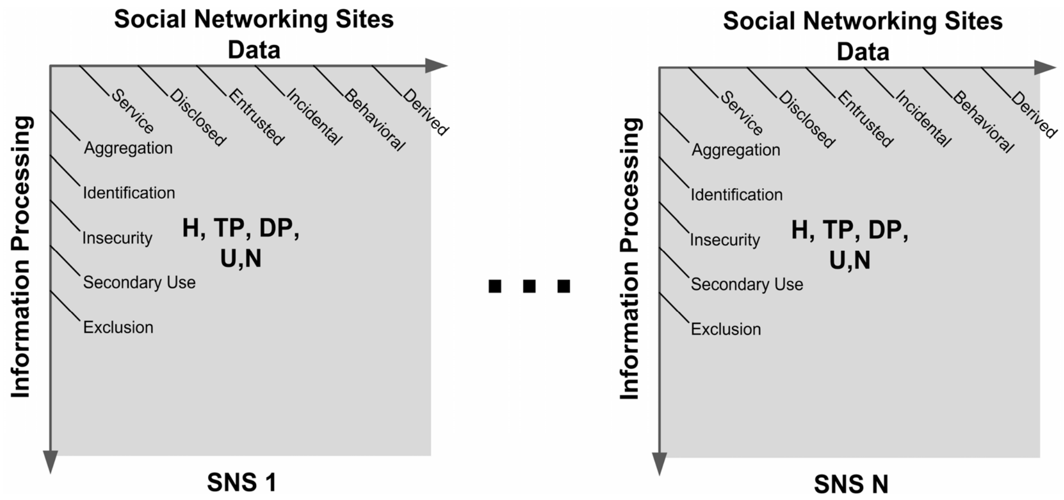

Figure 5. Part two: information processing.

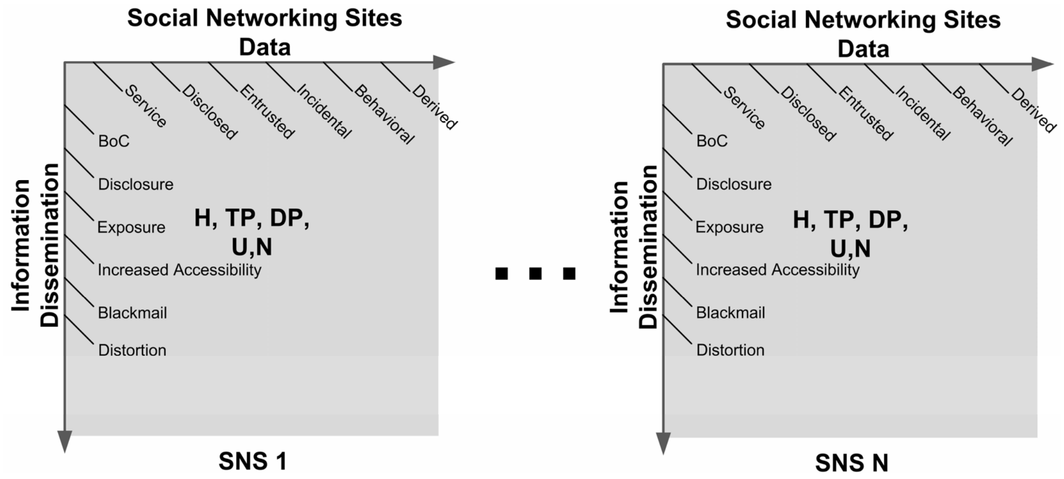

Figure 6. Part three: information dissemination.

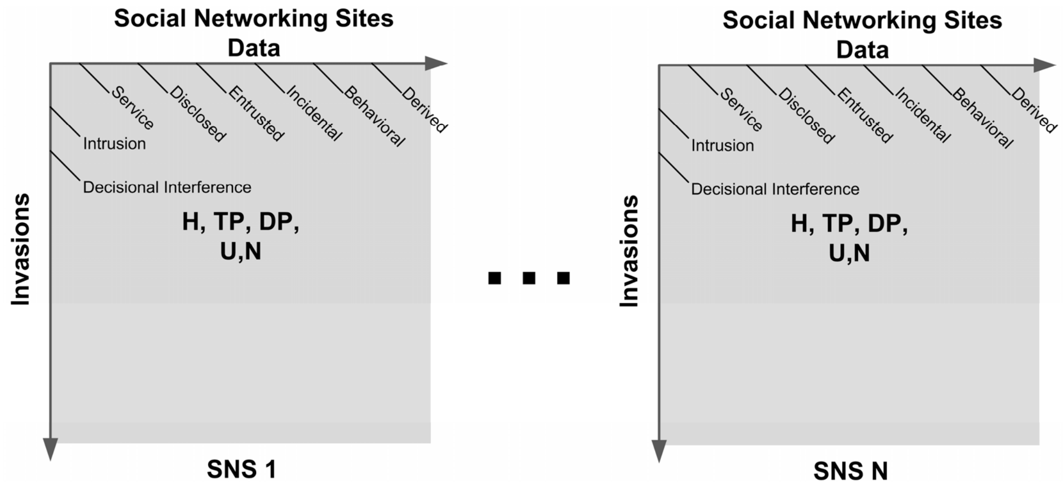

Figure 7. Part four: invasions.

essing, we are showing only the classification of Orkut and Orkut*, **Tables 6** and **7**.

There is a difference between the accessibility of user’s data when comparing Orkut and Orkut*. As re-

Table 3. Information collection classification.

IC	Ser.	Dis.	Ent.	Inc.	Beh.	Der.
Orkut	I	I	I	I	S	S
Facebook	I	I	I	I	S	S
Youtube	I	I	I	I	S	S
Twitter	I	I	I	I	S	S
LinkedIn	I	I	I	I	S	S
MySpace	I	I	I	I	S	S
Orkut*	**I**	**I**	**I**	**I**	**S**	**S**

Table 4. Information processing Orkut.

	User Data	Auto Generated Data
Aggregation	H	H, TP
Identification	N	N
Insecurity	N	N
Secondary Use	N	N
Exclusion	H, DP	H

Table 5. Information processing Orkut*.

	User Data	Auto Generated Data
Aggregation	H	H, TP
Identification	N	N
Insecurity	N	N
Secondary Use	N	N
Exclusion	H, DP	H

Table 6. Information dissemination Orkut.

	User Data	Auto Generated Data
BoC	N	N
Disclosure	N	N
Exposure	N	N
Increased Accessibility	H	H
Blackmail	N	N
Appropriation	N	N
Distortion	N	N

Table 7. Information dissemination Orkut*.

	User Data	Auto Generated Data
BoC	N	N
Disclosure	N	N
Exposure	N	N
Increased Accessibility	N	N
Blackmail	N	N
Appropriation	N	N
Distortion	N	N

gards to public data, Orkut may allow some user data to be viewed by the entire Web. This means that any user may see some data from a public profile if the data owner has granted such access. In Orkut*, the data I are still visible, but they are encrypted.

In the taxonomy, this increased ability to visualize the data was placed Increased Accessibility to the House because it is a feature allowed by the site. Another point that is not in the privacy policy of sites but must be taken into consideration is the Disclosure. This item consists of revelations about the user that may affect how the other members of the judge. Although not specified in the privacy policy on anything, it is common to see in social networking sites members posting such data. Therefore, an appropriate classification for this item would be {Disclosure, User Data} = U, which means that network users can make disclosure of user data.

Interestingly, almost all items were marked as N, not allowed. This happens because the dissemination of information can be viewed as something negative. Although it's possible to sites explicit leave something in this sense in its privacy policy, the use of this part of the taxonomy is ever more necessary in cases that privacy was invaded instead in privacy policy classification.

5.4. Invasions

This part of the taxonomy classifies invasion of privacy of individuals. As many users share data about their private life, this is an important part of the taxonomy.

About intrusion, no privacy policy says anything. But they reserve the right to share information if they are subpoenaed in court.

6. Discussion

This paper presented taxonomy for classifying the privacy policy of social networking sites and a privacy mechanism. The possibility of classifying both using a common taxonomy facilitates the evaluation in cases of invasion of privacy, because you can directly confront the SNSs privacy policy classification to the other, in which privacy was invaded.

By analyzing the privacy policies, we can see several things in common. Some subtle differences are made necessary by the type of service that each provides, for example, make the data visible to the entire Web or only for registered users. Although the privacy policy can increase the reliability of the user with the service, external entities can also reinforce this idea. Therefore, some social networks are part of the program of TRUSTe EU Safe Harbor.

We also presented and classified the privacy mechanism, which was implemented for Orkut. We called Orkut*, the privacy enhanced Orkut SNS powered with the privacy mechanism. We classified Orkut* and conclude the Orkut* offer a privacy enhance in terms of information dissemination.

REFERENCES

[1] A. L. Young and A. Quan-Haase, "Information Revelation and Internet Privacy Concerns on Social Network Sites: A Case Study of Facebook," *International Conference on Communities and Technologies*, University Park, New York, 2009, pp. 265-274.

[2] R. P. Botelho and S. D. Zorzo, "Privacify: Extending Privacy in Online Social Networking," *Lecture Notes in Informatik*, Berlin Informatik, Vol. 192, 2011, p. 432.

[3] B. Schneier, "A Taxonomy of Social Networking Data," *IEEE Security and Privacy*, Vol. 8, No. 4, 2010, p. 88. http://dx.doi.org/10.1109/MSP.2010.118

[4] D. M. Boyd and N. B. Ellison, "Social Network Sites: Definition, History, and Scholarship," *Journal of Computer-Mediated Communication*, Vol. 13, No. 1, 2008, pp. 210-230.

[5] L. Wu, *et al*., "Analysis of Social Networking Privacy Policies," *Proceedings of the* 2010 *EDBT/ICDT Workshops*, Lausanne, 2010.

[6] D. J. Solove, "A Taxonomy of Privacy," University of Pennsylvania Law Review, Vol. 154, No. 3, 2006, pp. 477-570. http://dx.doi.org/10.2307/40041279

[7] K. Barker, *et al*., "A Data Privacy Taxonomy," *Proceedings of the* 26*th British National Conference on Databases*: *Dataspace*: *The Final Frontier*, Vol. 5588, Springer-Verlag, Birmingham, 2009, pp. 42-54.

[8] A. Besmer, *et al*., "Social Applications: Exploring a More Secure Framework," *Symposium on Usable Privacy and Security*, Mountain View, ACM, California, 2009. http://dx.doi.org/10.1145/1572532.1572535

[9] J. Delgado, E. Rodríguez and S. Llorente, "User's Privacy in Applications Provided Through Social Networks," *Workshop on Social Media*, ACM SIGM, Firenze, 2010.