Privacy Enhancing Technologies (PET) and Web-Based Social Networks (WBSN)

Antoni Roig

1 IDT, Institute of Law and Technology,
Autonomous University of Barcelona
08193 Bellaterra (Barcelona) Spain
antoni.roig@uab.cat

Abstract. The technological threats to the right of privacy are not only limited to data bases. WBSN and pervasive computer, for instance, are two clear examples of other privacy risks. WBSN have an economic value, and more and more tools focus on WBSN users’ personal information. On the contrary, WBSN privacy is only a new research area. Internet communities are trust-based systems. Therefore, they need a privacy-respecting reputation system. Transparency tools should also allow individuals to check at any desired moment what personal data has been given to the data systems, and be able to alter or delete it. IT researchers usually consider privacy as a quantifiable attribute that can be negotiated and possibly exchanged by individuals in return for certain benefits. On the contrary, PET are necessary in WBSN. Thus, they cannot simply be individual options. Human rights, as public policies, should be preserved in the design of IT tools.

Key words: Web-Based Social Networks, Privacy Enhancing Technologies, privacy, privacy-respecting reputation system, social network analysis, semantic web.

1 Privacy and WBSN: legal framework and recommendations

E-privacy is often reduced in Europe to data protection. Nevertheless, the technological threats to the right of privacy are not only limited to data bases. WBSN and pervasive computer are two clear examples of other privacy risks. Certainly, the most significant privacy regulation is the European Data Protection Directive and the different implementations of the each Member State national law. Corresponding implements by national law. Indeed, the reports and studies of the national data protection agencies and its working groups, stated mostly art. 29 Data protection group, are the best legal framework.

One report of 2008 is perhaps up today the most relevant legal framework on WBSN and privacy (“Rome Memorandum” [1]). One key legal preliminary consideration is that there are only very few rules governing the publication of personal data at the initiative of private individuals. This is due to the fact that it had not been a major issue in the real world, and it has been only relevant on the Internet with WBSN. Another sociological key element is that a new generation of “digital natives” has appeared, and they seem more comfortable with publishing (sometimes intimate) details of their lives on the Internet. The Rome Memorandum recommendations to regulators are:

- Introducing the option of a right to pseudonymous use.
- Ensuring that service providers are honest and clear about what information is required for the basic service. Specific problems exist with consent of minors.
- Requiring data breach notification for social network services.
Another interesting document is the European Network and Information Security Agency Position Paper 1 [2]. Some of the recommendations are:

- WBSN should, where possible, use contextual information to educate people in ‘real-time’.
- Awareness-raising campaigns should also be directed at software developers to encourage security conscious development practices and corporate policy.
- The regulatory framework governing WBSN should be reviewed and, where necessary, revised:
  - What is the legal position on deletion of user generated content by service providers if it is classed as WBSN spam?
  - What is the legal position on image-tagging by third parties?
  - Who is responsible for security flaws resulting from user-generated markup or scripting?
  - How should privacy policies of embedded third party widgets be communicated to users?
  - What exactly constitutes personal data in a WBSN environment?
  - What is the legal position on profile-squatting?
  - Should the posting of certain classes of data by minors (location data) be made illegal?
- Users should be given accurate information on what is done with their data before and after the account closure.
- WBSN should be used in a controlled and open way (i.e. not banned or discouraged), with co-ordinated campaigns to educate children, teachers and parents.

2 Social Network Analysis (SNA) and WBSN

2.1 SNA

WBSN, or online communities [3], are growing with web 2.0 and with Mobile Social Software [4]. Indeed, WBSN have an economic value, and more and more tools focus on WBSN users’ personal information. This is known as SNA and it is complementary to the traditional questionnaire of the psychologists [5]. SNA is changing from initial mathematical graph theory and sociology to a more multidisciplinary research. Indeed, SNA will tend to include gradually social interaction and social reputation systems in a unified social computing framework [6]. Let’s see some SNA tools and we’ll see how far SNA is already gone.

The Organization Risk Analyzer (ORA) (www.casos.cs.cmu.edu/projects/ora) is a software program that computes social network. It also allows traditional node-link and advanced visualizations and user-editing of the meta-network data as well as provision of several other aids for advanced analysis, including error detection and what-if analysis using simulation tools [7]. SNA are mostly beneficial to sociologists or researchers in communities’ studies. However, some prototypes offer a personal
social network analysis service based on user’s reading and writing interest. This is the case of VisoLink, a user-centric SNA tool [8]. Social networking has also supported collaboration in computational Grids. Grids are complex systems that aggregate large amounts of distributed computational resources to perform large scale simulations and analysis by multiple research groups. Using social networking tools, Grid actors can discover partners to collaborate, potential providers and consumers [9]. In fact, more and more professionals are tending to collaborate in WBSN. For instance, in the field of medical science it is of paramount importance to share and circulate information about clinical cases and methodologies in the shortest time, as well as to create historical databases made available for future reference and analysis within the whole technical community [10]. On the other hand, the danger for privacy comes also from the involuntary information leakage in WBSN. For example, one research on a concrete WBSN has revealed that the first name of 72% of the accounts and the full name of 30% of the accounts could be easily inferred from the profiles by using a number of heuristics. The age of 15% of the account holders and at least one school attended by 42% of the holders could also be inferred [11].

2.2 Semantic Web and WBSN

In a WBSN we can only trust people we know. A network structure composed by trust statements linking individuals constitutes the basis for trusting people we do not know personally. This has been called “Web of Trust” [12]. We do not usually trust all the people who are trusted by the people we trust. It is not so easy. It would be useful if we could apply a trust metrics to compute how trustful a person is. Many ranking mechanisms and ways to compute trust relationships are considered at the moment. However, one might want to select the $n$ most trustworthy agents, while others would prefer all users with ranks above given thresholds. In any case, local group trust metrics, such as Advogato and Appleseed, offer interesting perspectives for diverse computing domains such as WBSN within the near future [13].

Why are there more and more Semantic web applications that address problems related to WBSN? One important reason is that the effort required to develop these applications has fallen in the past years: the standardization of the RDF and OWL ontology languages (2004) and SPARQL query language and protocol (2006) provide interoperability of semantic applications and services. Open source tools such as Sesame storage facility and programming tools like the Elmo API provide the common components of most Semantic Web applications [14]: for instance, we can mention Flink (http://flink.semanticweb.org) – a data collection for SNA –, or Openacademia (http://www.openacademia.org) – a publication metadata management. And, it is not only easier than in the past. Semantic web applications pretend to add the possibility of community-based ontology extraction from web pages. Thus, not only do we obtain the traditional network of ontology learning, but also a novel semantic network based on community relationships, a so-called emergent semantics: “it seems that ontologies are us: inseparable from the context of the community in which they are created and used” [14].

2.3 Privacy-Preserving Data Mining (P2DM) and WBSN

On the contrary, WBSN privacy is only a new research area. However, Privacy-preserving Data Mining (P2DM) is perhaps the exception. The main goal of P2DM is to avoid as much as possible the disclosure of private information about WBSN
members when analyzing users’ data for statistical purposes. General data protection can help, but WBSN are not data bases. Each data record in a data table is completely defined by the attribute values of a person, while a WBSN also contains relational data between individuals. In a WBSN, two users with the same public attribute values may still be distinguishable by their relationship with other users. Thus, not only has a privacy tool for WBSN to consider the attributes of the users but also the relationships between them [15]. These relationships could be determined using link discovery, and some efficient privacy-preserving link-discovery tools are also available [16]. Another possible solution is encryption for privacy-preserving collaborative WBSN [17]. This could help effective international security collaboration and personal information sharing. On the other hand, Peer-to-Peer sharing networks are also adopting some privacy-preserving tools [18].

3 Identity management, reputation systems and WBSN

Internet communities are trust-based systems. Therefore, they need a privacy-respecting reputation system as we will see later. The goal of a virtual community is to promote the enrolment of strangers and unknown users. But establishing the level of trust in the user requires ways to relate it with its trust value. We could use then real-world identities. As a result, privacy would be in danger because we would have personally identifiable information (PII). Therefore, more accurate reputation means usually less privacy protection.

3.1 Authentication without identification

Some authors try to have both privacy and trust. The identity management of users in an Internet community should be based, in their opinion, on recognition rather than authentication. As Seigneur says [19], in an authentication process we have:
- Enrolment: generally involves an administrator or human intervention.
- Triggering: someone clicks on a web link to a resource that requires authentication to be downloaded.
- Detective work: the main task is to verify that the entity’s claimed identity is the peer’s.
- Action: the identification is subsequently used in some ways.

On the other hand, the recognition consists of:
- Triggering (passive and active sense): the recognising user can trigger itself.
- Detective work: recognising the user.
- Upper-level action (optional): the outcome of the recognition is subsequently used in some ways.

The recognition process is an example of a more general replacement for authentication that does not necessarily bind an identity to the recognised identity. On the contrary, authentication is a recognition process that binds a real-world identity to the virtual identity. The possibility of recognising a user, analysing some of its attributes, is sufficient to establish trust in it based on past experience. One way of preserving both privacy and trust is using pseudonyms. Nevertheless, traffic analysis, data triangulation and data-mining can also associate a pseudonym with the real user. That is the reason why it is important that we provide multiple pseudonyms.

Technical solutions, such as a trust transfer, can be adopted then to avoid the misuse of multiple pseudonyms [19]. Another example is the EU-funded FP6 project PRIME (Privacy and Identity Management for Europe). Its approach uses “private
credentials” which enable proving one’s authorization (e.g., to be over 18 years old) without identifying the individual. They are derived from certificates issued on different pseudonyms of the same person, and they are neither linkable to each other nor to the issuance interaction. Only in the case of misuse, the user’s anonymity can be revoked [20].

Meanwhile, the anonymity analysis of supposed anonymous WBSN is also growing. A WBSN approach is adopted by assuming the attacker’s knowledge about users, based on the fact that they belong to such a network. Then, the performance of the anonymous WBSN in the context of this knowledge is evaluated. Furthermore, the analysis includes how errors in the information gained from the social network influence the correctness of the anonymity (and thus, the attacker’s confidence in her result) [21]. These authors notice that the overall anonymity is low and likely does not increase with the size of the social network. The positive aspect is that arbitrarily small errors in the profiles can lead to arbitrarily large errors in the anonymity probability distribution and hence point to the wrong subjects in the anonymity set.

### 3.2 Reputation Systems

**Reputation systems play an important role in Internet communities**

When people start in an Internet community, they usually have a pseudonym. Then, they have to gain reputation. In fact, it allows members of the community, to estimate other members’ behaviour before an interaction. Obviously, bad experiences are still possible: reputation is context-defined and subjective and someone might lie about another member or simply change his behaviour. Nevertheless, the virtual identity defined by the reputation is often trustworthy. The majority of the members of the community trust someone after confirming its reputation. We could imagine a trust metrics based on known people, Friend-of-A-Friend (FOAF); this was initially the core idea of the Trust Project available at http://trust.mindswap.org [22]. Nevertheless, it is not just a matter of number. Even with a low 0.01% of fraud, e-Bay is concerned because of the reputation of its e-trading system. The traditional legal dispute resolution –never excluded–, is not the way to solve millions of e-disputes. It simply does not allow the communities to maintain their reputation system. Online Dispute Resolution (ODR) will perhaps deserve a better service to Internet communities. In any case, a reputation system has to avoid disputes as much as possible.

**Privacy threats of current reputation systems**

Unfortunately the design of current reputation systems allows generating user profiles including all contexts the user has been involved in. This currently happens in electronic marketplace communities where time, frequency of participation, valuation of and interest in specific items, for instance, can be checked. Furthermore, trading partners usually have their pseudonym linked to a real name, so the profile is fully identified. Sometimes, only the provider is allowed to identify the partner, thus he has the possibility to inform about partners that have a bad reputation. But this centralised control is not always sure: it can be corrupted by a partner. In fact, most of mailing lists, newsgroups, discussion forums to role-playing and electronic marketplaces are implemented in a centralised way: a provider offers a technical system to the community. This is not so terrible. Even if we have a centralised internet community, we can protect privacy, while assuring a reputation system.
Another risk of reputation systems is due to the establishment of relationships of different types among users (e.g. friend of). In 2006, Facebook received the complaints of some users against the use of the News Feed feature, introduced to inform users with the latest personal information related to their online friends [23]. Facing an online petition to stop this service, signed by thousands of users, Facebook decided then to allow users to set some privacy preferences. In November 2007, another Facebook’s service concerned a lot of users: Beacon [24]. Beacon is part of the Facebook advertising system introduced to track users’ activities on web sites of Facebook partners. Such information was reported to users’ friends without the consent of the user itself. Some WBSN like Facebook or Videntity have then reacted giving to their users an optional mechanism: they can allow or not some users to access to their personal information (www.facebook.com, http://videntity.org).

More wide and flexible strategies are needed, making a user able to define his privacy policy or rules. Indeed, users should indicate which network participants are authorized to access his personal information, even though they are not directly connected through a relationship [25]. One option is through a client-side access control [26]. The requestor must provide the resource owner with a proof of being authorized to access the requested resource. Therefore, the privacy requirements established by WBSN users are preserved when enforcing access control. This privacy preserving access control can be even improved by a collaboration of selected nodes of the network [25]. The owner contacts only the nodes that satisfy its distribution rules. A node is invited to collaborate only if it satisfies the distribution rules of the other nodes taking part in the collaboration. Encryption and signature techniques are used to verify the correct enforcement of distribution rules. However, these techniques are not meant for general-purpose WBSNs like Facebook or MySpace, but for social networks used at the intranet level or by virtual organizations. With the same intention to protect the type and trust level of relationships, a public-key protocol has been described to achieve relationship protection with the advantage of not needing a central node. Besides, the new protocol avoids revealing the content of relationships to the resource requestor and substantially simplifies relationship revocation [27].

Access control is not the only way to preserve privacy in WBSN. More general privacy-enhancing reputation systems have to be adopted.

Privacy-respecting reputation systems

A more privacy-enhancing design of reputation systems is needed while keeping the trust provided to the members by the use of reputations. Even if trust is difficult to measure, a privacy-enhancing reputation system has to assure at least anonymity. And anonymity can be statistically calculated.

Let’s consider centralised reputation data bases [28]. Under some circumstances, users are allowed to rate other pseudonyms and these are updated to a central data base. A correct use of the pseudonym will be enough in many cases. The correct use of pseudonyms can increase easily trust in the reputation system. One of the problems that had appeared is the pseudonym changing without the transfer of the negative rates. Even in this case, technical solutions are also available. One possibility is to rate pseudonyms in different contexts: general or for specific issues, for instance. This could give a rate of one pseudonym more linked to each context the user is involved in. As long as additional information –number of user with a specific number of pseudonyms- is not provided, the anonymity is preserved. Another privacy-preserving reputation scheme has been proposed for a pseudonymous peer-to-peer (P2P) system [29]. Using e-cash for reputation points, the reputation of each user is closely related to his real identity rather than to his...
current pseudonym. Thus, a honest user can switch to a new pseudonym keeping his good reputation while a malicious user cannot erase his trail with a new pseudonym. This leads us to the next PET: transparency, context and purpose limitation.

4 Transparency, context and purpose limitation

We have just seen that authentication can adapt to a non-identification version, more flexible and less dangerous for privacy. Nevertheless, there is a deep change in the Privacy Enhancing Technologies (PET) related to the web 2.0, and above all online social networks: transparency is preferred to anonymity, authentication and other traditional protections. Old PET are still useful, but they do not give enough guaranty to new privacy threats.

Prior to an interaction, information on the interacting party should be made transparent. Transparency tools can provide clear visibility of the data flow, the privacy policy, present methods of data processing, offered services, used software, reputation of interaction partners, guarantees of trustworthiness and security of all data processing and also all present or possible vulnerabilities and security breaches. Although transparency tools alone are no panacea for maintaining the private sphere, the combination of transparency tools and user-controlled identity management systems yields viable functionality to empower users to protect their privacy [20].

One extreme possibility is to use TETs (Transparency Enhancing Technologies), so as to anticipate profiles that may be applied to a particular data subject. This concern personalized profiles as well as distributive or non-distributive group profiles, possibly constructed out of anonymous data. The central idea is to know the selection mechanisms (application of profiles) that may be applied. To be able to achieve this, the data subject needs access to additional external data sources. Then, based on this information, one can perform a counterprofiling.

More accepted is that transparency tools should allow individuals to check at any desired moment what personal data has been given to the data systems, and be able to alter or delete it. In PRIME, one of the main transparency tools is the “Data Track”, together with the function to check which personal data to disclose in a specific context. The information about potential interaction partners or the privacy policy can be shown before any disclosure of personal data. This information also allows users to ask data controllers later whether or not they have done what they have or to investigate potential risks related to former uses of the “Data Track”.

The value of transparency tools depends on how precise and understandable the information is. Standardization could help humans to understand and machines to interpret the information made transparent. Another challenge is that a transparency process can also be privacy-invasive. Therefore, data minimization with minimal disclosure of personal information is usually more effective than relying on “notice and choice” [20].

Context and purpose limitation attach to the identifiable data is also a new PET tool useful with web 2.0, the participatory web.

5 Conclusions

Privacy protection for pervasive computer can be a good analogy for PETs and WBSN. Privacy and self-presentation are compatible with automatic data disclosure, even if in ubiquitous computer, the awareness systems should be designed not to facilitate observation but to facilitate disclosure [30]. We can also consider WBSN
personalized privacy. In this case, the devices are the other “users” of a Personal Area Network (P.A.N.). A tool could analyze the requirements of potential users and automatically adapt the information visible according to the context and the individual privacy preferences of the user. This has already been designed for ubiquitous computer [31]. On the other hand, some analysis and simulations in pervasive computer are also adapted for social networks [32]. Even a general-purpose architecture for leveraging users’ mobile devices for measuring context, while maintaining the privacy of the users such as AnonySense could be adapted perhaps to WBSN [33].

It has been said on privacy in ubiquitous environments that we are witnessing a significant change: up to now, it was the role of the government to provide the framework for privacy protection. However, lately IT researchers tend to shift privacy protection into the hands of the individuals and to provide them with privacy protection mechanisms and tools. Furthermore, IT researchers usually consider privacy as a quantifiable attribute that can be negotiated and possibly exchanged by individuals in return for certain benefits [34]. On the contrary, PETs are necessary in WBSN. Thus, they cannot simply be individual options. Human rights, as principles, should be included in the design of IT tools. And this is not just a decision of industrial standards. It is above all a public policy to adopt. Privacy is perhaps the first right that can disappear if not protected at the first level of the design of technological tools, as a public policy. Other rights will soon follow the previous mentioned.

References

Privacy Enhancing Technologies (PET) and Web-Based Social Networks (WBSN)

27. Steinbrecher, S., Design Options for Privacy-Respecting Reputation Systems within Centralised Internet Communities, in IFIP International Federation for Information Processing, Security and Privacy in Dynamic Environments, eds. Fischer-Hübner, S,


