Expert Systems With Applications 62 (2016) 148–160

Contents lists available at ScienceDirect

Expert Systems With Applications

journal homepage: www.elsevier.com/locate/eswa

# Enforcing transparent access to private content in social networks by means of automatic sanitization

Alexandre Viejo, David Sánchez*

*UNESCO Chair in Data Privacy, Department of Computer Science and Mathematics, Universitat Rovira i Virgili, Avda. Països Catalans, 26, 43007 Tarragona, Spain*

ARTICLE INFO

*Article history:*
Received 11 January 2016
Accepted 13 June 2016
Available online 15 June 2016

*Keywords:*
Data publishing
Data protection
Social networks
Text sanitization
Privacy

ABSTRACT

Social networks have become an essential meeting point for millions of individuals willing to publish and consume huge quantities of heterogeneous information. Some studies have shown that the data published in these platforms may contain sensitive personal information and that external entities can gather and exploit this knowledge for their own benefit. Even though some methods to preserve the privacy of social networks users have been proposed, they generally apply rigid access control measures to the protected content and, even worse, they do not enable the users to understand which contents are sensitive. Last but not least, most of them require the collaboration of social network operators or they fail to provide a practical solution capable of working with well-known and already deployed social platforms. In this paper, we propose a new scheme that addresses all these issues. The new system is envisaged as an independent piece of software that does not depend on the social network in use and that can be transparently applied to most existing ones. According to a set of privacy requirements intuitively defined by the users of a social network, the proposed scheme is able to: (i) automatically detect sensitive data in users' publications; (ii) construct sanitized versions of such data; and (iii) provide privacy-preserving transparent access to sensitive contents by disclosing more or less information to readers according to their credentials toward the owner of the publications. We also study the applicability of the proposed system in general and illustrate its behavior in two case studies.



## 1. Introduction

A social network is a virtual environment powered by Web 2.0 technologies that enables users to publish and share all kinds of information and services with a global audience. Well-known platforms such as Facebook or Twitter have brought together more than 800 and 100 million active users respectively that *generate* and *consume* social contents (McMillan, 2011).

The Consumer Reports' (2010) State of the Net analysis (Consumer Reports National Research Center, 2010) stated that more than half of the users of social networks share private information about themselves online. This shows that the published content may contain sensitive personal data and, thus, it may represent a serious privacy threat (Velásquez, 2013). For example, some entities may exploit that knowledge to obtain benefits for their business (e.g. personalized spamming, phishing, etc.) (Zhang, Sun, Zhu, & Fang, 2010); recruiters may use it to hire or discard candidates (Snowdon, 2011); or it can even be used by regular people to perform bullying in the workplace or in the classroom (D'Arcy, 2011).

In recent years, social network users have been increasingly aware of that situation and studies have shown that their privacy concerns have negatively affected the way they use these applications (Staddon, Huffaker, Larking, & Sedley, 2012). Specifically, it has been reported that privacy-aware users spend less time posting and/or commenting social content. Due to the fact that an important part of the business model of the social network operators depends on the use of the social content generated by the users to attract advertisers (Crimes, 2012), the privacy concerns of the privacy-aware users represent a significant problem to the economic success of these platforms.

In order to minimize this issue, some social network operators have implemented limited privacy settings that allow users to decide who will have access to certain contents such as their profile attributes or their published messages. For example, Twitter allows their users to keep their messages (i.e., tweets) public (this is the default setting) or to protect them. Public tweets are visible to anyone, whether or not they have a Twitter account, while protected tweets are only visible to approved Twitter followers. On

* Corresponding author.
*E-mail addresses:* alexandre.viejo@urv.cat (A. Viejo), david.sanchez@urv.cat (D. Sánchez).

http://dx.doi.org/10.1016/j.eswa.2016.06.026
0957-4174/

the other hand, Facebook enables the owners of the social data to provide access to different groups of users (anyone, friends, friends of friends, co-workers, etc.) or even to single users.

Even though these privacy-preserving mechanisms represent a certain improvement with respect to the user's privacy, they have been criticized in the literature due to the following three essential problems: (i) these privacy settings are generally not sufficiently understood by the average users who seldom change the default configuration that generally makes most of the user information public (Bilton, 2010; Stern & Kumar, 2014; Van Eecke & Truyens, 2010); (ii) users are not informed about the privacy risks that their published data may cause and, therefore, they may find difficulties in defining effective privacy settings over their data (Wang, Nepali, & Nikolai, 2014); and (iii) these privacy settings, in any case, do not prevent social network operators from gathering sensitive user data and exploiting it to obtain economic benefits from advertisers or other entities (Crimes, 2012; Viejo, Castellà-Roca, & Rufián, 2013; Wilson, 2011).

Those problems have been already identified in the literature. Works such as Becker and Chen (2009), Talukder, Ouzzani, Elmagarmid, Elmeleegy, and Yakout (2010) and Wang et al. (2014) have provided solutions that measure the users' privacy exposure by means of their published profile attributes (e.g., address, political views, religious views, etc.) and warn them when their exposure level is too high; also, schemes such as Viejo et al. (2013) or Conti, Hasani, and Crispo (2011) have been designed to protect the sensitive user data from unauthorized entities, which include social network operators. Nevertheless, despite the efforts of the scientific community, there are two related issues that still have room for improvement: (i) there is not an effective mechanism that enables users to measure the degree of sensitivity of their textual publications (i.e., how dangerous is a certain tweet/timeline post from the privacy point of view?); and (ii) the privacy settings usually implemented in social networks offer a very *rigid* access control, that is, these methods manage the protected data as an indivisible element and, as a result, users either get full access to the whole protected data or they cannot obtain anything.

According to these points, new privacy-preserving mechanisms for social networks should be designed and, ideally, these new schemes should automatically warn the users about the privacy risks inherent to their textual publications and propose ways to reduce those hazards. Methods that offer access to contents should also be more flexible and transparent to the users. Finally, those systems should be deployed and managed by the users themselves to prevent the social network operators from gathering social data at will.

### 1.1. Previous work

As discussed earlier, the privacy settings implemented by social networks operators have been widely disqualified by the scientific community and some alternative approaches have been presented to protect the privacy of the users in a more effective way. We next review the most relevant ones.

Developing user privacy policies (as a contract that specifies who can access to a certain resource) for social networks and enforcing their application to ensure the proper protection of private data is a well-known approach used in works such as Aimeur (2010) , Carminati, Ferrari, Heatherly, Kantarcioglu, and Thuraisingham (2011), Cheek and Shehab (2012) and Dhia, Abdessalem, and Sozio (2012). A main shortcoming of this approach is that it requires the social network operators to implement such policies. Due to the fact that it is not clear which benefits would get an operator that implements those methods, it can be assumed that, for the moment, this approach is unlikely to be applied. Moreover, in any case, this approach does not prevent the social network operator from obtaining the protected data.

A more straightforward way to achieve this would be to create new privacy-enabled social networks specifically designed to apply strong privacy policies and to really preserve the privacy of their users. However, as stressed in Baden, Bender, Spring, and Bhattacharjee (2009), it is unrealistic to assume that a new brand social network can replace the well-known existing ones. In order to deal with this situation, the authors of this last paper present *Persona,* a social network integrated into Facebook as an application to which users log in through a Firefox extension. The users of Persona define a privacy policy that manages the access to their personal information. As a result, only the users with the appropriate access rights can get the protected data. Nevertheless, this tool is only a Facebook application that can be easily removed by the social network operator from the applications directory.

In spite of the low success probabilities of brand new social networks, there are some proposals in the literature that focus on designing new privacy-enabled social networks based on completely distributed architectures. Diaspora[1] is probably the clearest example of this approach; however, other systems such as Cutillo, Molva, and Strufe (2009), Nilizadeh, Jahid, Mittal, Borisov, and Kapadia (2012) and Vu, Aberer, Buchegger, and Datta (2009) can also be found in the literature. This kind of schemes allows users to install and manage their own personal web server to store all their data (e.g., photos, videos, etc.). Since each user controls her data server, she retains full ownership over the shared content and is not subjected to changing privacy policies and sell-outs to third parties (Vaughan-Nichols, 2010). The general idea of this approach is interesting but, as explained above, it is quite unlikely that those distributed social networks will attract enough users to become a real alternative.

The last approach considered in the literature is based on replacing the data to be protected by fake information that looks realistic in front of the social network operator. Under this approach, fake information is published in the social network while the real data is protected and stored in a way that the social network operator cannot obtain it. All the schemes in this category provide a software component that is in charge of transparently showing the real information when an authorized user browses the protected profile of another user. Note that the use of fake data that looks realistic is essential in those schemes due to the fact that the terms of service of well-known social networks usually state that it is forbidden to completely obfuscate the published personal data. In fact, Facebook has banned users who have violated those terms (Scoble, 2008). This behavior prevents straightforward solutions such as using cryptography to cipher text or attributes before publishing them.

Some researchers have followed this approach (Conti, et al., 2011; Luo, Xie, & Hengartner, 2009; Viejo et al., 2013) . The main difference between these schemes is the place where the real data is stored. In Luo et al. (2009), the real data is stored in an external centralized infrastructure that must be honest and always available. In Conti et al. (2011), the authors state that assuming the existence of a fully-honest and always-available third party server is unrealistic; therefore, the authors propose to locally store the real data in the computers of authorized friends. This behavior presents two main problems: (i) users are required to always connect to the social network by using the computer that locally stores the real data; and (ii) whenever a user modifies her protected information, it has to be individually sent to all the authorized friends for updating. Given that the society has already entered the age of ubiquitous connectivity, the first issue is a very

[1] http://joindiaspora.com (last accessed: January 2016).