# A semantic approach to security in social networks

Ana Ballester[1], Carles de-Haro[1], Francisco Jordan[2,1] and Helena Pujol[2]

[1] Universitat Politècnica de Catalunya, Departament d'Arquitectura de Computadors,
Campus Nord, Mòdul D6, Jordi Girona 1-3, E-08034 Barcelona, Spain
{anab, cdeharo, jordan}@ac.upc.edu
[2] Safelayer Secure Communications S.A.,
World Trade Center (Sud-4ª), Moll de Barcelona, s/n, E-08039 Barcelona, Spain
{jordan, helena.pujol}@safelayer.com

**Abstract.** With social networks getting highly popular and becoming a routine in most people's life, some security problems have arisen. These problems include issues like trust, interoperability or privacy. Social networks glean lots of information which cannot always be trusted because neither its source nor its integrity can be verified. Moreover, information is isolated in each social network, and the interoperation between sites for sharing purposes is not possible. Sharing information among users or social networks brings about a privacy problem, as a user should be able to define her own access control rules exactly as she wishes, and not based upon each site capabilities. Our proposal to solve these issues involves a Semantic Interoperability and Access Control layer (SIAC) combined with eXtended Web Of Trust (XWOT) ontology. SIAC will use identity ontologies, like FOAF and SIOC, to address interoperability, and semantic rule languages to solve the access control problem. XWOT, which is a PKI extension of the WOT ontology, may solve trustworthiness issues.

**Keywords:** social network, access control, trust, semantics, interoperability, ontology, PKI

## 1 Introduction

Content-sharing sites and social networks in particular have brought about revolutionary changes in the way users play a part in the Internet. Nowadays users no longer simply consume Internet resources, but they also create, publish and rank contents, leaving a trace of their digital activity, and giving away a great deal of their personal information.

For about twenty years, users have acted as mere information consumers, and trust issues have been solved by means of technologies like PKI, that allowed verifying websites and establishing SSL connections. However, for some time now, communications and web standards have evolved to give a rise to new collaborative web applications like wikis, social networks or calendars, which are open, user-friendly and, usually, free of charge. In fact, social applications where personal information and contents are shared have become so popular that millions of users

already have an account at sites like Facebook, MySpace, LinkedIn, Twitter or Orkut [1].

This huge success implies new concerns, not only related to the user's privacy, but also to information trustworthiness. In the first place, users build up their digital shadow and identity when they contribute to social applications with posts, tags, comments, etc., so they should be able to undoubtedly state which information comes from a reliable source or not. Thus, information origin and integrity could be verified.

Besides, as users keep leaving public traces of their personal and professional activity, they should be able to control who may access that information. This access control is nowadays regulated by each site's policy, forcing the user to understand each one of these policies, and to configure her privacy parameters in each one of the applications. Unfortunately, users are seldom strict when it comes to establishing and maintaining access policies, either because of unawareness, or lack of interest, or even application bugs.

Moreover, not only privacy policies are isolated in each site but also user's personal information. As the user is required to introduce her data in each of her accounts, information is often duplicated, incoherent and out-of-date. Therefore, there should be some interoperability solution in order to make the most of information that users have already entered.

In order to deal with these issues we propose to develop a Semantic Interoperability and Access Control layer (SIAC), which makes applications independent from data and from privacy policies, and empowers the user to take control over her own personal information.

The paper is organized as follows. Section 2 introduces the SIAC model. Section 3 outlines how Semantic Web standards are used to consolidate information from different sources, and how they may help social networks to interoperate. Section 4 extends semantic interoperability through a user-centric access control layer, and Section 5 discusses how to add trust to the information held by social networks. Finally, conclusions of our work can be found in Section 6.

## 2 Semantic Interoperability and Access Control layer

The SIAC layer handles the aforementioned concerns from a user-centric point of view, and answers both interoperability and security requirements, including privacy and trust. In particular, our proposal is based on the combination of semantic technologies, PKI and a policy driven access control mechanisms.

The SIAC layer is built upon the idea that all the information managed by social applications can be semantically expressed, in particular by ontologies like FOAF [2] or SIOC [3]. We also support this idea for the implementation of the access control layer, which will also use semantic languages, like SWRL or SPARQL, to express the access control rules that will drive the orchestration between social networks and users.

Thanks to this layered approach, data can remain distributed whereas access policies are centralized in a single access point, from where the user can easily manage them.

Our main contributions include the development of a prototype which is able to glean all the information of an entity distributed all over the social networks, allowing the user to have a global and centralized vision of all her public data. Another remarkable contribution is an extension of the WOT ontology [4] in order to support XML signatures.

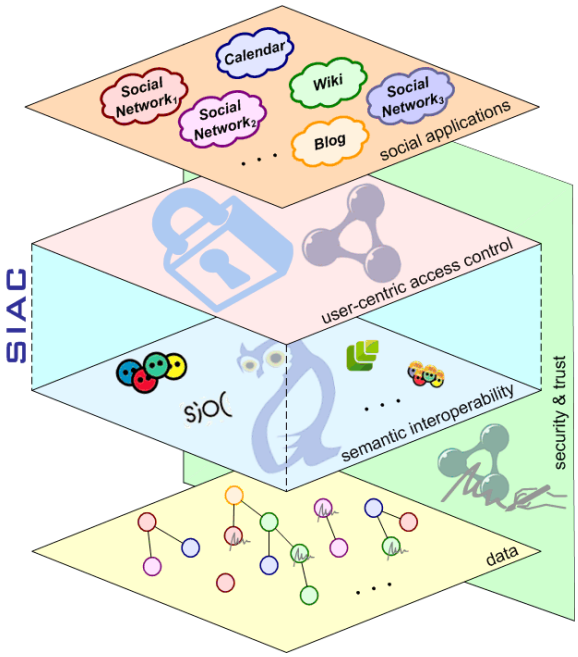


**Fig. 1.** Layered division for the social networks semantic interoperability. Social networks are connected to data repositories through a SIAC layer which is in turn divided into two other sublayers: one for the semantic access control, and another one for the ontologies that make possible semantic interoperability.

## 3 Using Semantic Web standards to consolidate and interoperate

User's information is stored differently in each social site. As a user may be registered in several sites, she has to manage her information in each one of the sites, causing the data to be scattered, duplicated and disorganized. If there was a way to represent the information in a common language, different social networks would be able to interoperate.

The *data layer* is composed of independent pieces of information which may be stored in different repositories and comply with diverse information structures. Independently of how this information is expressed, we should be able to represent it through ontologies and Semantic Web standards [5], which are the appropriate solution to achieve interoperability among social networks. An ontology is a "formal

representation of a set of concepts within a domain and the relationships between those concepts"; therefore, we can use common sets of vocabulary to describe identity, relationships between individuals and social networking activities.

In order to represent the personal information attributes stored by social networks, we propose the use of FOAF and SIOC, as well as other related ontologies. The FOAF ontology is a result from the Friend Of A Friend Project, which is probably one of the most successful initiatives within the Semantic Web. FOAF models basic identity attributes and relationships between individuals, either persons or organizations, so it allows describing the identity and contacts of an entity in the context of social networks on the Internet through a set of RDF statements. On the other hand, the Semantically-Interlinked Online Communities initiative aims to enable the integration of social network information by means of an ontology, as well. The SIOC ontology complements the FOAF ontology, allowing the representation of social application concepts in RDF.

In order to overcome the identity information dispersion issue, we have analysed the data models used in several social networks so as to match them with identity ontologies. For those attributes that couldn't be directly matched to the FOAF specification, we searched for other popular ontologies that covered other particular sets of attributes, for example, those related to geolocation or résumé.

There are basically two access mechanisms that popular social networks offer to retrieve personal data from outside their platforms: access to attributes through an API (generally through a REST interface, as in Facebook or Orkut), and access to an automatically generated FOAF profile (as in Proofile, LiveJournal, Vox, CrazyLife, Origo, e-Cademy and Hi5). In this case, the social network publishes an instance of the user's FOAF profile, although they don't always strictly stick to the FOAF ontology definitions, or they interpret FOAF classes and properties in a different way.

The percentage of identity information that can be directly expressed by the FOAF ontology is 42% for Facebook's API, F8, and 38% for Google's one, OpenSocial. The rest of useful information can be represented by other identity ontologies like DOAC [6], Dublin Core [7] or vCard [8], among others.

Should all social networks provide a semantic interface, they could be able to interoperate and retrieve information from any other network, without the need to store duplicated information. Then, for example, the user could fill in her career information in only one social network, and allow other sites to retrieve this information, according to the access control policies explained in section 4.

Thus, our user-centric approach lets the users control where their information is stored, and who may access it. As a first step to achieve this, our FOAF Manager prototype acts as an aggregator of identity attributes and demonstrates that semantic technologies are suited to fulfil this purpose.

### 3.1 The FOAF Manager prototype

The FOAF Manager prototype retrieves the user's identity attributes that are stored in different social networks, and consolidates the selected information in a single FOAF file. The user must specify which sites the application should retrieve data from, and if requested, authenticate at those sites.

Once the various profiles are loaded in the FOAF Manager, the application shows the information in RDF syntax, and assists the user in the edition, management, signature and publication of the FOAF files. It also includes the functionality to merge different FOAF files into a single one automatically, which can be very useful when combining information relating to friends from different social networks.

Thanks to the FOAF Manager application, a user is able to examine the identity information about her, scattered all over the social networks, and may detect out-of-date attributes or even private information which shouldn't be shared. This is the first stage of a more complete application that should allow the user to set privacy policies on her identity attributes.

## 4 Using ontologies to enhance access control

Social networks usually provide ad hoc and inflexible access control mechanisms. In general, a user profile is divided in blocks of information that can be completely public or private, and the granularity of these blocks is defined by the site, not by the user. Moreover, privacy requirements that are easily stated in natural language are highly complicated to articulate through access control policy languages. Furthermore, policies are distributed in every social site, so the user is forced to configure many different policies independently, adapting her desires to those rules permitted by the site, which may not always satisfy her expectations.

Some previous studies [9] [10] [11] have attempted to improve access control mechanisms in social networks. However, these proposals neither consider the use of semantic technologies nor endorse a user-centric vision.

The SIAC model makes data and access policies independent from applications. Should the identity information be represented in a semantic language at the *semantic interoperability layer*, particularly as a set of RDF resources, semantic policy languages could be used to control access to every resource described in the user's profile (e.g., friends, jobs, contact data, etc.). Semantic access control languages already allow writing policies in which the action, the resource and the decision are described by semantic statements. Thus, the user could decide which resource (and up to which granularity level) would be controlled by a particular rule. For instance, a user may grant access to her personal phone number to all of her colleagues, only to her family members, etc. Alternatively, access policies could be expressed using the SPARQL semantic query language or the Semantic Web Rule Language (SWRL), as well.

The SIAC *user-centric access control layer* would be used to rule the access to every user's profile throughout all social networks. This single privacy policy would be easier to maintain and tune, and automatic tools could be developed in order to check rules consistency and recommend modifications. As a result, defining a single user-centric access policy would increase user's privacy and information security.

All social sites could be interconnected through the SIAC layer enforced by the user's defined policy, which will specify the information accessible by each network. Whenever a user registers in a new site, she should specify where her policy is published, in order to allow the social network to access the information stored in

other social networks. Ideally, this access policy could be stored in a unique repository, which could be any of the social networks the user is registered in, or any other user's personal repository. Therefore, allowing social networks to gather information from other applications, according to user's policies, would decrease information duplication and obsolescence.

Since the information will remain distributed, the user will have to define two kinds of rules in her access control policy. On one hand, the current rules already used by social networks will still be in use, allowing a user to block access to some pieces of her profiles to other users. On the other hand, new rules will be necessary to manage the access among the different networks. Both other users and social networks will be understood as resources.

## 5 Using PKI to increase trust

Former sections introduce a solution for the problems of dispersion and access control to the information, but we still have to deal with guaranteeing the proof of origin and integrity of data, which is a transversal issue. Traditionally, these problems have been solved with the use of PKI technologies. Nevertheless, in the SIAC proposal we will not just use classic PKI, but we will adapt it to fit our semantic approach, as well, as there are many issues that can't be solved from the traditional electronic signature perspective, but could be addressed with a new semantic signature concept.

The first aspect to be noted is that traditional signature ensures the integrity of information, interpreted as data or bytes. However, with the use of semantic technologies bytes are no longer the unit of information: concepts are. From a semantic perspective one could sign concepts, even allowing for changes in their syntactic representation if those changes would not affect its meaning. Then, from a semantic perspective, "bytes integrity" gives way to "information integrity", where information is more than just a bunch of bytes.

In classical signature, authentication and no-repudiation issues are solved by using digital certificates, which guarantee signer's identity, as her public key is invariably linked to a set of her personal data. Nevertheless, with the unstoppable advance of user centric systems, as well as the view that there are multiple identities that identify the same entity, and that an identity is a collection of flexible identity information attributes, the classic concept of digital certificates falls short to address these new trends.

A semantic signature approach would also allow including interesting context information on the circumstances the signature was performed under, and would effectively contribute to view and understand the meaning of the information long time after it was created.

Ideally, we would like to use a semantic signature like the one we described above, but this is not a trivial issue yet. In fact, there are some studies about RDF graphs semantic signatures [12] [13], but they still lack of a general consent.

The approach given in this paper proposes an intermediate solution in which the signature will still be a classic PKI signature, using the XAdES standard [14], but will be stated in a semantic format. This intermediate solution, called eXtended WOT

ontology (XWOT), is based on the Web Of Trust ontology, which was originally intended to sign and encrypt RDF information with PGP and GPG keys. Our extended version of the ontology supports PKI concepts and XML signatures. For this purpose, we added two new classes: XMLEndorsement, which contains a detached XML signature, and Certificate, that represents the X.509 certificate which was used to sign.

We have also defined three new properties: signedBy, which bonds the signature (or XMLEndorsement) with the Certificate involved; hasCertificate, which associates a Certificate to a User, and containsKey, which associates a public key to a particular certificate.

Thus, although we still don't pre-process RDF documents with canonicalization algorithms, we can somehow ensure information's integrity and origin's identity.

This XWOT signature could be used to sign user's attributes in social networks. Besides a user being able to sign a piece of her own data, a more interesting example consists of a third party signing some user's attribute, thus assuring the trustworthiness of that particular information. Then, educational institutions would issue signed RDF blocks stating that the user had finished a course, an administration would state the user's address, or an enterprise would recommend a former employee. These signed pieces of RDF would increase the trust in the distributed data stored by the different social networks.

In the particular case of relationships, current social networks expect both entities to accept that this relationship exists; with the use of signatures this would not change, because both parties would have to sign that a relationship exists, but this process would be performed only once and would be verifiable by all the social networks without the need to confirm it again whenever the user creates an account in another site.

## 6 Conclusions

The Semantic Interoperability and Access Control proposal implements a user-centric model, where social applications are independent from identity data as well as from privacy policies.

By means of Semantic Web standards (like RDF, OWL, SPARQL and SRWL) and the definition of ontologies, identity information can be represented in a machine-processable common syntax. This is the first step to make the most of the information that users have already stored in social networks, in order to share it and avoid multiplying the sites where the same attribute must be uploaded and maintained. The FOAF Manager prototype already collects identity attributes from various sources, and translates them to FOAF and other identity ontologies in a single global view.

On top of this semantic interoperability layer, every user should be able to define a single access control policy that all social networks should observe. Our proposal lets the user control the access to her information by the use of semantic rule languages (SPARQL or SWRL), being able to decide the level of granularity of each access control rule and benefiting from the expressiveness and power of semantic relations. The access control layer could be used to specify which users have access to which

information, as well as which social networks have access to that information, because from the user's global point of view both are stated as RDF resources. Additionally, specific tools could be developed in order to check policy consistency and coherence.

Finally, in order to implement the transversal concept of trust to the SIAC model, we propose to combine PKI technologies and the WOT ontology, defining a new ontology called XWOT. Using this ontology, information's origin and integrity could be asserted and verified, and trusted third parties could issue pieces of reliable identity information that a user might include in her personal profile.

Thanks to the SIAC model, social applications could share, trust and reuse identity information, and it would be easier for users to control it.

**Acknowledgments**. This research has been supported by Safelayer Secure Communications and the Centre for the Development of Industrial Technology (CDTI) of Spain, within the framework of the Segur@ project, reference CENIT-2007 2004 of the CENIT Programme (part of the INGENIO 2010 initiative) [15].

## References

1. "Social Networks: Facebook Takes Over Top Spot, Twitter Climbs", http://blog.compete.com/2009/02/09/facebook-myspace-twitter-social-network/
2. "The Friend of a Friend (FOAF) project", http://www.foaf-project.org/
3. "Semantically-Interlinked Online Communities Project", http://sioc-project.org/
4. "WOT (Web of Trust) Schema", http://xmlns.com/wot/0.1/
5. "W3C Semantic Web Activity", World Wide Web Consortium, http://www.w3.org/2001/sw/
6. "DOAC: Description of a Career", http://ramonantonio.net/doac/0.1/
7. "Dublin Core Metadata Initiative", http://dublincore.org/
8. "An ontology for vCards", http://www.w3.org/2006/vcard/ns
9. A. Tootoonchian, K.K. Gollu, S. Saroiu, Y. Ganjali, and A. Wolman, "Lockr: social access control for web 2.0," Proceedings of the first workshop on Online social networks, ACM New York, NY, USA, 2008, pp. 43-48.
10. M. Hart, R. Johnson, and A. Stent, "More content-less control: Access control in the Web 2.0," Proceedings of the IEEE Web 2.0 Privacy and Security Workshop.
11. B. Carminati, E. Ferrari, and A. Perego, "Rule-based access control for social networks," Lecture Notes in Computer Science, vol. 4278, 2006, p. 1734.
12. J. J. Carroll, "Signing RDF graphs", Digital Media Systems Laboratory, July 2003.
13. J. J. Carroll, C. Bizer, P. Hayes, P. Stickler, "Named Graphs, Provenance and Trust", HP technical report, 2004.
14. "XML Advanced Electronic Signatures (XAdES)," ETSI TS 101 903 V1.3.2, Technical Specification, vol. 3, 2006.
15. "Proyecto CENIT SEGUR@ Seguridad y Confianza en la Sociedad de la Información", http://www.cenitsegura.com