Online user's registration respecting privacy

Aude Plateaux, Patrick Lacharme, Christophe Rosenberger, Kumar Murty

Online user's registration respecting privacy

Aude Plateaux[1),2)] , Patrick Lacharme[1)]
[1)] Greyc Laboratory, ENSICAEN - CNRS, UCBN,
14000 Caen, FRANCE
[2)] BULL SAS – Payment Systems & PKI Entity,
78340 Les Clayes-Sous-Bois, FRANCE
aude.plateaux@ensicaen.fr, patrick.lacharme@ensicaen.fr

Christophe Rosenberger[1)] , Kumar Murty[3)]
[3)] Department of Mathematics, 40 St. George Street,
Toronto, CANADA
christophe.rosenberger@ensicaen.fr,
murty@math.toronto.edu

***Abstract* — The explosion of Internet concerns a lot of domains from e-commerce to social networks. Various data, including personal data, are exchanged and stored in large databases or possibly in cloud storage, by numerous websites or online merchants. However, these service providers do not necessarily propose a clear policy on their personal data collection and storage. At the same time, users of the Internet need to understand how these recordings are being carried out. In this paper, a new solution is proposed for the registration to a website which helps the user to control his/her data. This solution is easy to use, user centric and privacy compliant.**

Keywords: privacy, online recording, user centric architectures

I. INTRODUCTION

In 10 years, the number of computers connected has increased by 10,000 to reach 10 million. Today, the Dot-com bubble exploded represents over 37 million users. Internet offers more and more possibilities and services for various applications, from social networks to e-commerce. In most of cases, Internet users must provide various personal data in order to access a service of a web site. These latter are possibly stored without real control or time limitation. Moreover, most requested information is only used for market research or advertising purposes. The proliferation of personal data gives even greater importance to the issue of privacy on the Internet.

Each country has its own legislation in terms of privacy protection. The United States protects young Internet users by COPPA [8], consumers with the Gramm-Leach-Bliley act [9] and medical information by HIPAA [4]. The European Union has the Directive 95/46 Data Protection [2]. Finally, Canada uses PIPEDA act for the protection of personal data [17].

Resolution 45/95 [1] adopted by the General Assembly of the United Nations provides, among other things, the principle of data security. This principle ensures that appropriate measures must be taken to protect computerized data against natural and human risks. These risks can be unauthorized access, misuse of data and computer viruses. However, we must add to the data security the requirements in terms of privacy. Three principles for this protection are developed:

- Data sensitivity principle: Handled personal data are considered as sensitive, requiring a de-centralized structure for data storage.
- Data sovereignty principle: Personal data belong to an individual, with a control and consent on how these data are used and for what purpose.
- Data minimization principle: Personal data disclosure should be limited to adequate, relevant and non-excessive data. It includes anonymity and unlinkability of data.

These two last principles are detailed in the paper and Deswarte Gambs concerning the French national identity card [22], and the PRIME project [21]. This latter attempts to develop a framework for identity management in order to protect the privacy. In addition, in November, 2011, the European Commission examined ways to strengthen the data minimization principle. In addition, the European Commission has started the PrimeLife project [23]. Its purpose is to allow users to identify web sites which do not respect the users' personal information. Consequently, in contrast to Mark Zuckerberg's comments, privacy is not an outdated concept.

The access to web service begins by a first phase of registration. Users should provide various information to the service provider, SP. These access conditions generally begin by a valid email address, a user name and a password, associated with the web site. Additional personal information are often expected or required, including personal address, phone number, date of birth and favorite occupations and leisures. In addition, the user usually needs to answer to a question at the end of the questionnaire. He/She must check or uncheck the invisible box: "I agree to receive email with offers selected for me by the SP and its partner." However, to enable the SP and its partners to select user's preferences, these providers use client's personal information and follow the user's comings and goings on the site. Moreover, in many cases, the user's consent is not explicitly requested. For example, French sites automatically uncheck this box (opt-in consent), whereas English sites prefer opt-out consent. Thus, if the user does not refuse these offers by unchecking the box, advertising e-mails are automatically sent to him.

After the enrollment, the second step is the client's authentication with the couple login-password previously chosen. This phase is also invasive in terms of privacy protection. Indeed, the client authenticates in first, without proof of the SP authenticity and without conviction to deal with this SP. Thus, the SP registers the client personal information in its database and can follow the client's navigation for each connection.

Use of personal data for promotional offers or other various advertising messages is a strong problem for privacy. Moreover, users are rarely aware of the issues in terms of privacy when they disclose personal information. Personal data

are "information relating to a natural person who is or can be identified, directly or indirectly" [5] and must be protected. It is therefore necessary to modify the recording protocols for existing websites. The sensitivity and the scores of authorized information to be disclosed must be limited, and users should be alerted on the use of these data.

Contribution. Most of publications focus on authentication protocols on Internet or on a specific technology improving privacy. This paper proposes a new enrollment solution with a service provider on Internet. The architecture provides an ease of recording using a form and an analyzer of access conditions, enhancing users' privacy. This approach respects the data sovereignty, minimization and sensitivity principles.

Organization. This paper begins with a brief state of the art concerning authentication, anonymity protocols and specific web technologies. The new architecture is presented in Section III and Section IV. This solution is analyzed in Section V.

II. PRIVACY ENHANCING TECHNOLOGIES

In this section, we review some privacy concepts, used in our architecture. To our knowledge, very few enrollment architectures have been proposed to date. The known architectures generally allow to assure one privacy property but never all important principles.

A. Specific web technologies

Web-specific technologies exist to protect users' personal data. The P3P platform [12], for instance, allows websites to express their practices in terms of privacy. The advantage is this format can be quickly obtained and interpreted by user agents. These latter give the possibility to be informed of site practices to clients. Users avoid the difficult task of knowledge concerning the privacy policies of all visited sites. Unfortunately, although the P3P is consistent with the laws and self-regulatory program, it does not provide a mechanism to check site compliance with its policies. Similarly, in order to help consumers, IBM offers a risk and security management infrastructure [13]. It allows users to manage their personal information and control the privacy policies of online sites.

Numerous existing technologies concern social networks. Diaspora [24], Peerson [25] and Safebook [26] are examples of these architectures which allow to protect user's data. Diaspora allows users to set up their own servers to host data without need to bounce communications. However, Diaspora provides no support for encryption. Peerson is a peer-to-peer infrastructure including encryption protocol in order to users keep control of their data. Moreover, the users can use this social network without Internet access. Finally, Safebook does not precise enough its policies. Jahid et al. in [27] make a critical study concerning privacy performance of these infrastructures. Thus, they propose a new decentralized architecture for social networks. Unfortunately, these architectures are often complex for ordinary persons. Moreover, to our knowledge, no online registration architecture enables to protect clients' personal information by disclosing a minimum amount and respecting the data sensibility. Thus, in the two following subsections, we describe the necessary and existing knowledge for our architecture.

B. Identity and authentication

An online service generally begins with an authentication and a secure connection between the client and the SP website, using a protocol such as SSL/TLS [28, 29]. This protocol involves the client trusting in the SP and is aware of known published browser attacks [30, 31, 32]. In the proposed architecture, the authentication is ensured by strong authentication or zero knowledge authentication, whereas data access control is based on the entity's identity, a list of access rights and the sensitivity of the protected information. Multi-factors authentication generally uses a password and a smart card. In some cases, an additional factor is added. This may be a biometric factor, as in [15]. However, zero-knowledge authentication is easy to realize on a standard computer. These protocols can prove a given value, for example the identity, without revealing it. The ISO/IEC 97/98 gives explicit mechanisms using the zero-knowledge principle [3]. Classical zero knowledge authentication protocols are the Fiat-Shamir protocol [10] and the Schnorr protocol [20].

C. Anonymity and pseudonymity

Anonymity and pseudonymity preserve the users' privacy during a connection to commercial sites. The FPR class (Privacy) of the Common Criteria establishes security criteria for information systems [14]. This class is composed of four requirements: anonymity, pseudonymity, unlinkability and non-observability. These two first services, as well as the data minimization principle are also discussed by Pfitzmann and Hansen [18] and by Cameron [7]. These requirements can be used for identity management preserving privacy. The anonymity ensures the user can access to information without revealing his/her identity, whereas the pseudonymity requires the user is responsible for such use. There is a persistent assumed identity, namely the pseudonym. A single person may have many pseudonyms.

Consequently, in exceptional cases, the identity can be retrieved by authorized persons. Many anonymity and pseudonymity systems exist. For example, the German act [6] requires their service providers to offer anonymous or pseudonymous payment services for their users. Another famous example of anonymity system is the onion rounting for anonymous communications [19, 11].

III. ENROLLMENT ARCHITECTURE BETWEEN CLIENT AND SP

A. Overview

The proposed architecture for registration online service provider is centered on user privacy. In using this solution, the users could easily enroll in revealing a minimum of personal and sensitive data. The architecture ensures the client does not provide personal information if he/she has not obtained the SP authentication. Thus, the client reveals neither his/her data, nor his/her desires. The client only provides the necessary, relevant and adequate information for the transaction in order to assure the minimization, sensibility and sovereignty principles. The architecture is summarized in Fig. 1. This new online registration architecture involves four main actors: the user or client, the service provider SP, the client's bank and an advice part or counselor, directly connected to the user.

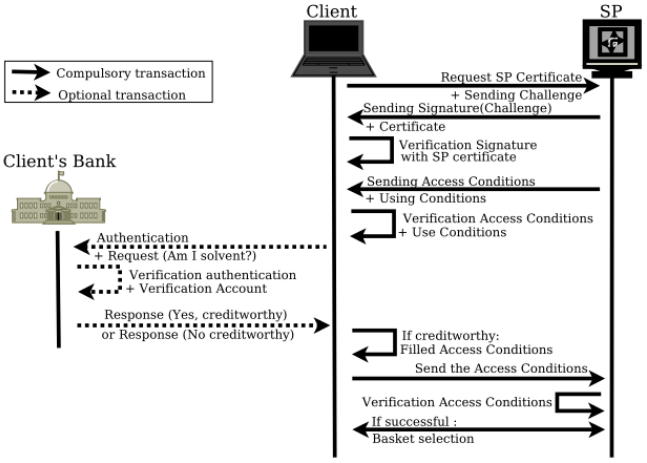

Figure 1. The new enrollment architecture when the client is not registered.

The SP has a key pair whose public key is certified by the payment scheme(s) it has contractualized with. This certificate contains the following compulsory parameters:

- Its name which proves to the client that there is a SP with a valid certificate authority;
- Parameters describing the payment scheme recognizing the SP and allowing to secure the future payment (AmEx, VISA, MasterCard...);
- Specific certificates parameters:
  - Designation of the certification authority;
  - Validation period;
  - Name of the public key holder;
  - Identification of the signature algorithm ;
  - Identification of the encryption algorithm;
  - Identification of the encryption and signature algorithms;
- A public key of an asymmetric keys pair: $K_{public}(SP)$;
- The name of certification authority;
- The digital signature value. In our case, it is the signature of a challenge sent by the Client.

As far as the counselor is concerned, it helps the user to be aware about dangers of the data disclosure and provides tools to protect the personal information. Thus, this actor is directly related to the user, facilitates the client's process and helps him to bring into focus of the data sensitivity. The client can decide to use these available options. The counselor can play different roles: analyzer of access conditions, analyzer of conditions of usage, provider of proof of knowledge, identity manager or electronic strong-box. The counselor is software installed on the client's computer.

B. The enrollment architecture

Enrollment takes place in different ways with more or less options. Thus, the more options are used, the more client's privacy protection is assured. This section explains the main transactions between the SP and the client. The various and optional possibilities given by the counselor are detailed in the next section. Firstly, the client browses on the website and asks the SP to be authenticated, using a traditional challenge/response protocol. The client sends a random challenge to the SP. The SP signs the challenge using its private key in order to prove ownership of the certificate and this signed challenge is sent to the client. The client verifies the SP signature using the SP public key. After the SP authentication, the client takes into consideration services and access conditions. If he/she accepts them, he/she sends the fulfilled access conditions to the SP, with a couple of login/password for futures connections. If sensitive or personal data are required, a mathematical proof of knowledge is only delivered for the SP. These data include the date of birth and the card number.

In case the client does not know the balance of his/her account, or if he/she prefers to have a confirmation from his/her bank, he/she may request a pre-authorization. In a single operation, the client authenticates to his/her bank and requests a pre-authorization for a given amount. The authentication can be done by a traditional login/password or through a CAP reader [16]. If the authentication is successful, the bank positively responds to the request by checking the client's account balance for a given amount. Thus, in case the client is not creditworthy, the client avoids providing information to SP.

As the next step, after having the SP authentication and being sure to deal with the SP and to be solvent, the client authenticates to the SP thanks to his/her login/password. However, if the SP does not expect specific access conditions and simply the authentication without disclosure, the client can use a zero-knownledge authentication protocol. We do not detail this possibility in the article. Indeed, actually the SP gives advantage to client records for statistics and adverts. However, in order to respect the data sovereignty principle, the client is the one to decide on sending of the access conditions.

Finally, if the client agrees, he/she sends the access conditions to the SP. This latter verifies this document and his/her proof of knowledge replacing the disclosure of date of birth. If the access data are correct, the payment phase between actors can process.

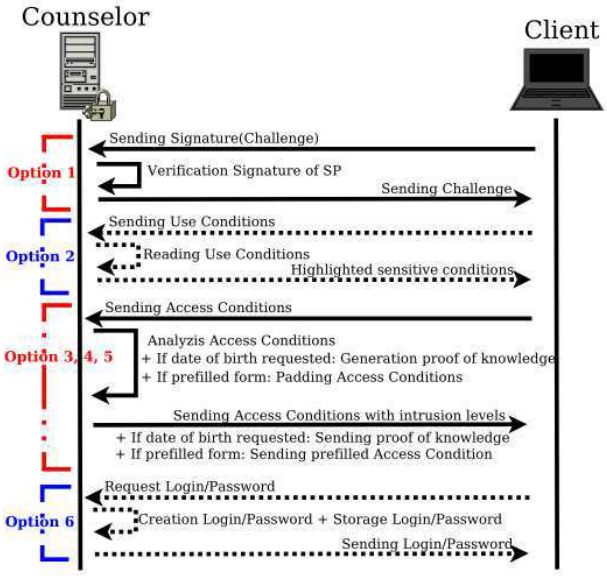

Figure 2. The available transactions with the counselor.

IV. ENHANCING PRIVACY WITH THE COUNSELOR

The counselor is an application on the client's computer, managing and analyzing personal data exchange. The client decides to use the available options of the counselor, as summarized in Fig. 2.

Firstly, the counselor can play the role of signature verifier. In this case, the client easily sends the challenge to the SP for its authentication. Then, the counselor checks the SP signature of this challenge.

In the same way, as described previously, the client provides a proof of knowledge of personal data, as the date of birth. The counselor takes responsibility for this proof.

Secondly, the client can request an analysis of the conditions of usage. The counselor simply read and highlights important terms of the contract in terms of privacy protection. This actor may also analyze access requirements. The counselor then adds to each required information one levels of intrusion in terms of privacy. These two actions allow the client to manage personal information and to minimize their disclosures.

Moreover, the client can pre-fill a form containing data which can be revealed. If this form exists, the counselor uses it to pre-fill the access conditions. The padding of this form has been guided by the counselor in order to know the level of data sensitivity. With the knowledge of the data sensitivity and authorized conditions, the choice is left to the client to include other information. For instance, the form can contain information with conditions such as:

- An email address can be useful in the case of password loss;
- The mailing address is only relevant in the case of home delivery. This requirement is that this address must be provided at the end of the process;
- The card number is stored only if the client requires. Otherwise, it is a real sensitive data;
- The client's age if the access requires a minimum age. The date of birth is not provided except when the client requires. Otherwise, we prefer use a proof of knowledge;
- An option is checked if the client accepts to receive advertisements or "newsletters".

Thus, in the case where the SP requests a condition present in this form, it is not necessary to ask the client's consent. This method simplifies the connection to the website for the client. However, if a requirement is not contained in the form, or if it contains a requirement (as in the mailing address), the client's consent is automatically requested.

One solution is to create a common form for all SP with adequate, relevant and non-sensitive information. Moreover, it is possible to ensure the authenticity of the information containing in the form. This latter must be signed by a certification authority. The client may, at any time, update this document and return for signature to the certification authority. This form is stored by the counselor in an electronic strong-box. Thus, once the access conditions pre-filled by the counselor, the client checks these conditions (and eventually completes them).

Finally, in order to avoid the problem of forgotten passwords, it is common to use the same couple. However, the multiple uses of the same couple cause trouble for associability of data and identity theft. Thus, the client can also use the counselor as an identity manager. This actor generates one such pair for each SP. Then, in the same way as the form, these couples are stored in a secure digital strong-box. Obviously, in a future connection to the SP, the couple allows the client not to register again.

V. ANALYSIS OF THE PROPOSED PROTOCOL

The security of this architecture is linked to the possession of a certificate by the SP, the different authentication algorithm, as well as secure channels used for each transaction. The counselor is able to securely generate and store different couples of login/password for each SP. Thus, the user avoids the data correlation problem in the case of a corrupted database, as well as theft of identity.

In addition, the proposed solution is entirely focused on the client. The access conditions and conditions of use are analyzed by the counselor. Then, a form can be completed by the client with the help of the counselor. A degree of intrusion in terms of privacy is also indicated during the analysis and the padding of the form. Thus, the data sensitivity principle is respected.

The principle of minimization is also assured. The data are only disclosed after the SP authentication and proof of creditworthiness. In addition, only relevant and adequate data are disclosed and a proof of knowledge is provided instead of the date of birth.

Finally, the client's consent is required at all steps. He/She has the last word concerning his/her data. Consequently, the data sovereignty principle is also respected.

VI. CONCLUSION

The large amount of personal information stored by service providers in their databases requires protection with particular attention given to privacy issues. The users' data are possibly re-used for statistics and adverts and rarely considered in terms of privacy protection.

This new architecture uses a pre-filled form, a reorganization of transactions and known cryptographic tools, such as certificates, proof of knowledge and authentication protocols. Furthermore, the counselor is able to play several roles: signature verifier, analyzer of conditions, provider of proof of knowledge, identity manager or manager of electronic strong-box. These options help users in their approach and make them aware of the data sensitivity. Thus, the proposition offers a secure and easy-to-use solution respecting the fundamental principles in terms of user's privacy.

VII. Acknowledgment

The authors would like to thank Vincent MARIE for his participation during the implementation, as well as the BULL company and the ANRT association for their financial support.

References

[1] Guidelines for the regulation of computerized personal data files. Adopted by General Assembly, resolution 45/95, December 1990.

[2] Directive 95/46/EC of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data. Official Journal L, 281(23/11), 0031-0050.

[3] ISO/IEC 97/98-5 Information technology - Security techniques - Entity authentication - Part 5: Mechanisms using Zero-knowledge techniques. 2009.

[4] A. Act. Health insurance portability and accountability act of 1996. Public Law, 104:191, 1996.

[5] Assemblée nationale et Sénat. Loi 78-17: Lois Informatiques et Libertés, 1978.

[6] L. Bygrave. Germanys teleservices data protection Act. Privacy Law & Policy Reporter, 5:53–55, 1998.

[7] K. Cameron. The laws of identity. Microsoft Corp.

[8] Federal Trade Commission. Coppa - Children's Online Privacy Protection Act. 2002.

[9] J.C. Cuaresma. The Gramm-Leach-Bliley act. Berkeley Tech. LJ, 17:497, 2002.

[10] A. Fiat and A. Shamir. How to prove yourself: Practical solutions to identification and signature problems. In Advances in CryptologyCrypto86, pages 186–194. Springer, 1987.

[11] D. Goldschlag, M. Reed, and P. Syverson. Onion routing. Communications of the ACM, 42(2):39–41, 1999.

[12] W3C Working Group. The platform for privacy preferences specification, November, 13 2006.

[13] IBM. Tivoli security products, 1996.

[14] P.K. Infrastructure and T.P. Profile. Common criteria for information technology security evaluation. 2002.

[15] S. Krawczyk and A. Jain. Securing electronic medical records using biometric authentication. In Audio-and Video-Based Biometric Person Authentication, pages 1110–1119. Springer, 2005.

[16] MasterCard. Chip authentication program functional architecture, September, 2004.

[17] Privacy Commissioner of Canada. PIPEDA - The Personal Information Protection and Electronic Documents Act. Office of the Privacy Commissioner of Canada, 2010.

[18] A. Pfitzmann and M. Hansen. Anonymity, unlinkability, unobservability, pseudonymity, and identity management - a consolidated proposal for terminology. Technical Report, 2008. v0.31.

[19] M.G. Reed, P.F. Syverson, and D.M. Goldschlag. Anonymous connections and onion routing. Selected Areas in Communications, IEEE Journal on, 16(4):482–494, 1998.

[20] C.P. Schnorr. Efficient signature generation by smart cards. Journal of cryptology, 4(3):161–174, 1991., A Treatise on Electricity and Magnetism, 3rd ed., vol. 2. Oxford: Clarendon, 1892, pp.68–73.

[21] PRIME project, Privacy and Identity Management for Europe, May 2008.

[22] Y. Deswarte and S. Gambs. Towards a privacy-preserving national identity card. Data Privacy Management and Autonomous Spontaneous Security, pages 48–64, 2010.

[23] Vimercati, G.N. and Paraboschi, S. and Pedrini, E. and Preiss, F-S. and Raggett, D. and Samarati, P. and Trabelsi, S. and Verdicchio, M., Primelife policy language, 2009.

[24] Project Diaspora, 2010.

[25] Buchegger, S., Schiöberg, D., Vu, L., Datta, A.: Peerson: P2P social network-ing: early experiences and insights. In: Proceedings of the Second ACM EuroSysWorkshop on Social Network Systems, ACM (2009) 46–52.

[26] Cutillo, L., Molva, R., Strufe, T.: Safebook: Feasibility of transitive cooperation for privacy on a decentralized social network. In: World of Wireless, Mobile and Multimedia Networks & Workshops, 2009. WoWMoM 2009. IEEE International Symposium on a, IEEE (2009) 1–6.

[27] Jahid, S., Nilizadeh, S., Mittal, P., Borisov, N., Kapadia, A.: Decent - a decentralized architecture for enforcing privacy in online social networks. December, 2011.

[28] A. Freier, P. Kocher, and P. Karlton. RFC 6101: The Secure Sockets Layer (SSL) protocol version 3.0, 2011.

[29] T. Dierks. RFC 5246: The Transport Layer Security (TLS) protocol version 1.2, 2008.

[30] D. Wagner and B. Schneier. Analysis of the ssl 3.0 protocol. In The Second USENIX Workshop on Electronic Commerce Proceedings, pages 29–40, 1996.

[31] E. Gabrilovich and A. Gontmakher. The homograph attack. Communications of the ACM, 45(2):128, 2002.

[32] O.Aciicmez, W. Schindler, and C¸ .K. Ko¸c. Improving Brumley and Boneh timing attack on unprotected ssl implementations. In Proceedings of the 12th ACM conference on Computer and communications security, pages 139–146. ACM, 2005.