A Framework for Privacy-Preserving E-learning

Esma AÏMEUR, Hicham HAGE and Flavien Serge MANI ONANA
Département d'informatique et de recherche opérationnelle
Université de Montréal
{aimeur, hagehich, manionaf}@iro.umontreal.ca

Abstract. E-learning systems have made considerable progress within the last few years. Nonetheless, the issue of learner privacy has been practically ignored. The security of E-learning systems offers some privacy protection, but remains unsatisfactory on several levels. In this work, we corroborate the need for privacy in E-learning systems. In particular, we introduce a framework for privacy preserving E-learning to provide the learner with the possibility of combining different levels of Privacy and *Tracking* to satisfy his personal privacy concerns. This allows the learner to perform learning activities and to prove his achievements (such as with anonymous transcripts and anonymous degrees) without exposing various aspects of his private data. In addition, we introduce the *Blind Digital Certificate*, a digital certificate that does not reveal the learner's identity. Finally, we report on the implementation and validation of our approach in the context of an E-testing system.

1 Introduction

E-learning emerged over 20 years ago. At first, it consisted solely of text, like a book on a screen, and was ineffective and unpopular with learners. Today, E-learning has become richer with multimedia content and more interactive. In a typical E-learning system, other than the Tutor and Learner interfaces, there are many components collaborating in order to analyze the learner's skills, develop and deliver proper training material, and evaluate the learning process. Nonetheless, for simplicity, we group these components into four major components: the Tutor Environment, the Learner Environment, Data Storage and the System Manager (Fig. 1). The Tutor Environment is composed of tools and functionalities that can be grouped into three major parts: Content Authoring to create learning objects and assessments or to import/export learning material from other systems. Course

Please use the following format when citing this chapter:

Aïmeur, E., Hage, H. and Mani Onana, F. S., 2007, in IFIP International Federation for Information Processing, Volume 238, Trust Management, eds. Etalle, S., Marsh, S., (Boston: Springer), pp. 223–238.

Management offers class management tools such as electronic grade books, and splitting the class for group work. Finally, Learner Tracking allows the tutor to track learner (or group) activities and performance, and to offer personalized feedback. The Learner Environment on the other hand offers the learner a Learning Environment to perform the learning tasks assigned by the tutor, a set of Productivity Tools such as to track his progress and performance, search tools, and a forum. In addition, Evaluation and Assessment tools allow the learner to take un-graded practice tests and quizzes, and to actually take a graded assessment.

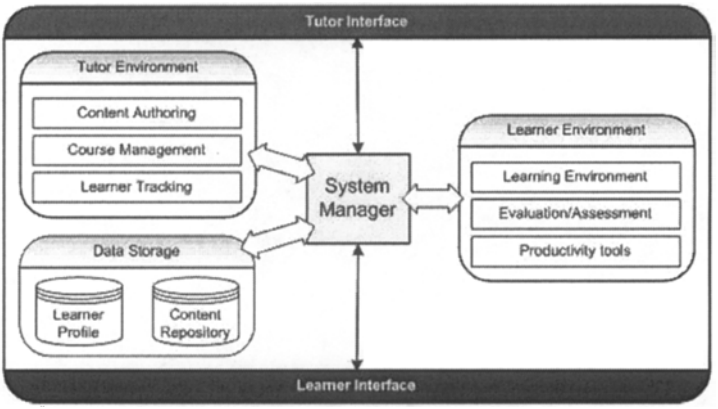

Fig. 1. Generic E-learning architecture

The Data Storage components contain all the necessary data and information. The Learner Profile stores all the relevant information about the learner such as identification data (name, age, gender, etc.), learning style, preferences and the courses the learner has passed/failed. On the other hand, the Content Repository contains the learning objects and assessments. The System Manager is a set of tools and protocols to handle communication and access privileges. For example the System Manager assures that a learner does not have access to the tutor environment and vice versa. Moreover, the System Manager arbitrates the access to the Data Storage so as to verify that each module or component is accessing the proper data.

One of the main advantages of E-learning is its adaptability to the learner's specific needs and preferences. But in order to do so, the E-learning systems must collect large amounts of information about the learner [1], thus violating his *privacy*, which is the claim of individuals to determine what information about themselves is known to others, as well as when and how it is used [2]. The security aspects of E-learning systems do offer some privacy protection; nonetheless it remains unsatisfactory on several levels. Other than the case of Head-in-the-sand privacy (by which the learner wants to keep secret his ignorance even from himself), learners might need to keep private different parts of their profile for different reasons. For example, a learner who is following a professional training course, for competitive reasons, would rather keep his identity hidden; yet, he wouldn't mind leaving a trace

of his activities in the E-learning system. On the other hand, a secret agent would rather take the training course for a top-secret mission without revealing his identity and without even leaving a trace that someone took this training. Thus, in order to satisfy various privacy needs, we adapt the levels of Privacy and the levels of Tracking introduced in [3-4] to the context of E-learning. In particular, learners are able to receive anonymous transcripts and anonymous degrees such as to prove their accomplishments and achievements to third entities (employers, other E-learning systems, etc.) without compromising their private data. Moreover, in order for the learner to prove that he is the rightful owner of the anonymous transcript or degree, we introduce the concept of *Blind Digital Certificates*, a digital certificate that does not reveal the learner's identity. Although issuing anonymous credentials and certificates is not a new idea, *Blind Digital Certificates* are designed for the specific structure of the e-learning environment. We are aware that not everybody will embrace our wish for privacy. Nevertheless, as many would agree, we consider privacy to be a fundamental human right: it is not negotiable! This is why we introduce Privacy-Preserving E-learning as an alternative to standard E-learning. Of course, the final choice belongs to each learner. As a proof of concept, we use public-key cryptography as well as digital signatures to implement and validate an E-testing system (the Evaluation/Assessment component in Fig. 1) along with the Privacy-Preserving Processes to satisfy the various privacy and tracking levels. In this E testing setting, depending on the selected privacy level, some of the learner profile's data is encrypted, and the remaining data is stored in the clear. Moreover, based on the learner's learning objectives, the E-testing system signs and authenticates his achievements.

The paper is organized as follows: Section 2 highlights the existing literature on privacy in E-learning and offers some preliminaries on Pseudonymous and Anonymous Credentials. Section 3 raises privacy issues in E-learning and our proposed framework to solve these issues. In Section 4, we introduce Privacy-Preserving E-learning. Section 5 details the implementation and validation of our approach in the context of E-testing. Section 6 offers a discussion of further issues to consider and Section 7 concludes the paper and offers pointers to future works.

2 Related work

2.1 Related work on privacy in E-learning

Although learner privacy is barely addressed within E-learning systems, there were concerns raised with regards to security. There exists literature, such as [5], on how to achieve two key security requirements: *confidentiality* and *integrity*, which provide a certain level of privacy. **Integrity** guarantees that the data is not maliciously or accidentally tampered with or modified: for example, when the learner submits his test, he requires the guarantee that his test answers are not modified after his submission. **Confidentiality** assures that the data and information is kept secret and private and is disclosed only to the authorized person(s): for

example, test scores must be accessible only to the appropriate tutor. The confidentiality of the information is considered at two different stages: while it is being transmitted to/from the E-learning system, and when it is stored within the E-learning system. In the first case, the data can be encrypted using Public Key Encryption such that only the appropriate receiver can read the data. In the second case, the use of access control mechanisms [6] can be employed to restrict access to the data. Access control cannot totally guarantee the privacy of the learner: first of all, it does not protect against a *super user* with full access privileges. Second, the learner has no control on which information about him is being gathered by the E-learning system. Although Privacy Policies have been provided for this purpose [7], they cannot restrict unwanted access to the data. One cannot preserve the learner's privacy without first identifying the issues and then defining the proper framework to solve these issues (Section 3.1 and 3.2).

2.2 Pseudonymous and Anonymous Credentials

Certificate Authorities (CA) are trusted entities whose central responsibility is certifying the authenticity of entities (persons or organizations) and their public keys. More precisely, an entity certificate issued and signed by a CA acts as proof that the legitimate public key is associated with the entity. Usually, the CA makes the decision to issue a digital certificate based on evidence or knowledge obtained in verifying the identity of the owner. In the context of privacy-preserving systems, the CA cannot be used to protect user private data and transactions. Therefore, new approaches are considered.

In 1985, Chaum [8] introduced the concept of pseudonymous credentials to protect individual privacy. More precisely, the resulting system enables users to engage in several anonymous and untraceable electronic transactions with organizations. Two years later, the implementation of this concept was proposed by Chaum and Evertse [9]. However, it was not suitable in practice because it relied on the existence of a semi-trusted third party participating in all communications.

In 2000, Brand [10] used several properties of Chaum's original concept to introduce a privacy-enhanced certificate system. Here, the system consists of two entities (Organizations and Users) and two protocols (Issue and Show). Unfortunately, Brand's approach is also limited for practical implementations. For instance, every Brand's credential is unique, thus it can be showed only once; otherwise, transactions by the same user could be linked. To overcome this limitation, the system needs to be extended by introducing recertification or batch issuing mechanisms [10].

Another implementation of Chaum's proposal is the credential system proposed by Camenisch and Lysyanskaya [11], which is based on previous work by Lysyanskaya et al. [12]. Here, users first register with the root pseudonym authority before using the system. Thus, users are unable to build up several parallel identities and they can be traced in case of fraudulent transactions. Users are limited to at most one pseudonym per organization. Each credential is related to a single attribute and

an expiry date. Moreover, users are able to choose which statement to prove about an attribute, such as choosing to prove that the value of attribute "age" is greater than 18. While considered an interesting implementation of the concept of pseudonyms and credentials, Brand's solution has the drawback of being based on zero knowledge proofs, thus the system is difficult to implement in environments with limited resources.

Although the previous general solutions for anonymity, pseudonyms and credentials can be used to solve issues related to user privacy in various domains, we aim to use the specific structure of an E-learning setting in order to seek more efficient solutions. Therefore, we introduce the concept of Blind Digital Certificates, to enable privacy-enhanced access to E-learning data.

3 Privacy in E-learning

3.1 Issues to solve

Privacy is nearly absent in current E-learning systems. Only primitive forms of privacy are offered in some platforms, such as not allowing the tutor any access to information about auto-evaluations performed by the learners. Nonetheless, the tutor has access to virtually all the remaining information including who are the students, what parts of the course they referred to, how many times and for how long, all the messages in the forums, and all the information about the quizzes and tests the learner took in his course. There are many reasons why a learner would like to keep his information private. We group these reasons under two main categories: *Competitive* and *Personal*. In the **Competitive** context, the learner requires his privacy due to competitive considerations. Consider, for example, a prominent politician taking a course to increase his knowledge in a certain domain, which will give him an advantage over his opponents. Other than for protecting himself from any prejudice from the part of the tutor, he has the right and interest in keeping this fact hidden, and his performance results private, from public knowledge and scrutiny, especially from his opponents. As another example, consider a company that uses E-learning for employee training purposes. If competitors have knowledge of the training and the performance of the employees, it could seriously affect the competitiveness of the company and its reputation, especially if the employees performed poorly. Finally, in the case of a military training E-learning system, just knowing that secret agents performed a specific training (such as desert or jungle survival techniques) could jeopardize their mission objectives. In the **Personal** context, the learner requires his privacy due to personal considerations. For example, he may wish to protect himself from a biased tutor. The bias of the tutor might stem from prejudice or stereotyping, based on a previous encounter with the learner, or even from personal reasons. Another reason a learner would prefer to keep his privacy is the increased pressure and stress due to performance anxiety; a learner might feel more comfortable and relaxed knowing the tutor will not know how he performed in the test.

In addition, there are issues to consider with regards to the learner's educational goal. An employee must be able to prove to his manager that he completed the training successfully without exposing his privacy. Moreover, a learner must be able to prove that he finished the prerequisites for a certain course, to assert that he has the required degree to obtain a job, and he should be able to pursue higher education while maintaining his privacy.

3.2 Framework for solving the issues

Our task is to provide an architecture for E-learning systems in which privacy issues can be addressed. With this in mind, we consider the following components of the learner's data, which are of interest from a privacy point of view.

- **The identity (id):** refers to information that makes it possible to determine physically who the learner is (or at least to seriously circumscribe the possibilities). This includes data such as his name, address, and student id number.
- **The demographic profile (dp):** refers to demographic characteristics of the learner, such as age, gender, weight, race, ethnic origin, language, etc.
- **The learning profile (lp):** refers to information such as the learner's qualifications, his learning style, interests, goal and aspirations.
- **The course history (ch):** lists the courses the learner has followed in the past, and their respective information such as the learner's activities within the course and his final grade.
- **The current courses (cc):** lists the courses in which the learner is currently registered and those he is attending, as well as the courses' respective information such as the learner's activities within the course.

These elements constitute the Learner Profile, L = (id, dp, lp, ch, cc). Moreover, we define, in this context, a learner's activity within a course as being any act involving one of the course's tools or resources. For example, an activity might involve the posting of a message in the forum, using one of the course's learning objects, or even taking a quiz or a test.

The above elements constitute the personal information on which we base our privacy framework for E-learning systems. Since different learners prefer different degrees of privacy, we adapt the four levels of privacy introduced in [3] to the context of E-learning.

1. **No Privacy:** the learner doesn't wish, or doesn't care to keep private any of his information. He does not mind the compilation of a dossier about him that consists of his identity, demographic information as well as his learning history.
2. **Soft Privacy:** the learner wants to keep his identity and demographic profile private, but he does not mind if the tutor has access to his learning profile, course history and current courses.

3. **Hard Privacy:** the learner wants to keep his identity, demographic profile, course history and learning profile private, but he does not mind if his current courses are known.
4. **Full Privacy:** the learner wants to keep secret every component of his personal data.

Another dimension to consider, which is independent of the personal data listed above, is the tracking of learners within a course. Even under soft, hard or full-privacy constraints, some learners might not want the tutor to know their activities and navigation within the system. Thus, we introduce the following terminology, inspired by [4], to account for the levels of tracking that different learners might accept.

1. **Strong Tracking:** the system can relate the activities performed within all the courses to the specific learner, even though that learner may be anonymous. In this case, the system can track the same learner u and his access to courses c_1, c_2 ... c_n.
2. **Average Tracking:** the system can relate the activities within a course to the same learner u, but cannot relate them to other activities within other courses. In this case, the system can relate the activity of u_1 in c_1, of u_2 in c_2 ... and of u_n in c_n, but cannot link u_1 to u_2 to ... u_n.
3. **Weak Tracking:** in this case, although the system recognizes the learner u as a regular visitor, it cannot link him to a course nor trace his activities.
4. **No Tracking:** in this case, the system cannot even recognize the learner u as a recurring user of the system.

For example, a learner, Alice, is using a privacy-aware E-learning system to take the following courses: CSC101 and CSC102. In the case of **Strong Tracking**, Alice creates a pseudonym, A, and uses it to access and perform the learning activities in CSC101 and CSC102. In the case of **Average Tracking**, Alice creates two pseudonyms, A1 and A2, one for each course, such that the system cannot relate A1 and A2 to Alice, nor to each other. Hence, whenever Alice needs to access and perform the learning activities in CSC101 or CSC102, she uses respectively A1 or A2. In the case of **Weak Tracking**, the system only records that Alice was logged in, but leaves no trace of her activity (nor identity). And, in the case of **No Tracking**, the system cannot even trace that Alice was logged in at all. Selecting No Tracking is similar to using a guest account to access a demo of the E-learning system. If the learner requires a proof of his achievements (Section 4.1), he must select at least Weak Tracking.

4 Privacy-Preserving E-learning

Now that the privacy issues are defined and the framework to solve these issues is set, we present our solution: Privacy-Preserving E-learning. We first introduce the

tools we use to protect learner privacy, and then we present our solution, which utilizes Public Key Cryptosystems (PKCs). In the remainder of this work, we use the following notation: E is an Encryption Algorithm which computes the ciphertext $c = E_{pk}(m)$ of a message m, given a public key pk; and D is a Decryption Algorithm which computes the cleartext message $m = D_{sk}(c) = D_{sk}(E_{pk}(m))$, back from the ciphertext c, given the secret key sk.

4.1 Anonymous Transcripts, Anonymous Degree and Blind Digital Certificate

There are many situations in which the learner will require some official document from the E-learning system to prove his achievements (to a third party or the E-learning system itself). Among other such documents are his transcripts as well as his degrees obtained within the E-learning system. For privacy purposes, these documents must remain anonymous, while being able to prove the credentials of the learner:

- **Anonymous transcripts:** the grades attributed for exams and assignments are grouped in the same document to form the learner's transcript, which remains an anonymous transcript since the E-learning system cannot identify the learner.
- **Anonymous degree:** similarly, an anonymous degree is a degree delivered to the learner by the E-learning system, such that the learner can prove that he earned this degree without revealing his identity.

In order to deliver an anonymous transcript or degree, the E-learning system uses the blind digital certificate as the learner's identifier, and uses its own private key to sign the anonymous transcript or anonymous degree. It is important to note that the anonymity of the degree and transcript does not increase the risk of *consensual impersonation* (where the actual learner asks someone else to take the course or perform the learning task in his stead), since it is an existing issue in traditional E-learning.

Moreover, in the context of privacy-preserving E-learning, the learner needs to prove to a third party (an employer, another E-learning system, etc.) that he is the rightful owner of a number of anonymous documents. Recall that a *Digital Certificate* is a certificate that uses a digital signature to bind a public key to an entity's (a person's or an organization's) identification data, which includes information such as name, address, etc. The main objective of the certificate is to prove that a public key belongs to a certain entity. Since the learner wishes to protect his privacy, a conventional digital certificate cannot be used. Therefore, we introduce the concept of *blind digital certificates*. We define a **blind digital certificate** as a certificate that binds together the learner's public key with the *encrypted* components of his profile (depending on the level of privacy). In particular, the learner's identity and demographic data will always be encrypted. For example, if $L = (id, dp, lp, ch, cc)$ is the learner's profile, and (pk, sk) is his public/private key-pair, then the corresponding blind digital certificate in the case of full privacy is the digital signature apposed by the CA on $[pk, E_{pk}(id), E_{pk}(dp), E_{pk}(lp), E_{pk}(ch), E_{pk}(cc)]$, where

$E_{pk}(x)$ is the public key encryption (Section 2.2) of x using the public key pk. Similarly, the corresponding blind digital certificate in the case of soft privacy is the digital signature apposed by the CA on [pk, $E_{pk}(id)$, $E_{pk}(dp)$, lp, ch, cc]. Please take note that instead of having only one id, the learner could create several identifiers id_1,..., id_n, if he wishes to prove his achievements to n entities, so that each pseudonym or identifier is used for only one entity. This is greatly inspired from Chaum's seminal fight against Big Brother [8], with his proposition of giving different pseudonyms to each organization with which he does business, so that it is impossible for organizations to link records from different sources and then generate dossiers on him.

4.2 Architecture

Fig. 2 illustrates a Privacy-Preserving E-learning architecture. Compared to the architecture in Fig. 1, the difference, although very subtle, is clear: since the System Manager controls data flow between the Tutor/Learner Environments and the Data Storage, it is only logical to include the Privacy-Preserving Processes into the System Manager. Moreover, the Learner Profile was split into two parts: the Basic Learner Profile and the Anonymous Learner Profile. The Basic Learner Profile functions as in any privacy-deprived E-learning environment, storing the data in the clear. On the other hand, the Anonymous Learner Profile stores the Learner's profile and information privately. In this case, a tutor has access only to the data in the Basic Learner Profile. Even if the tutor can access the Anonymous Learner Profile, he cannot decrypt it and view the learner's data. Similarly, the Privacy-Preserving System Manager performs a Privacy-Preserving Process, depending on the learner's privacy preference, before updating the learner's profile. The next section details the Privacy-Preserving Processes in the case of Soft Privacy.

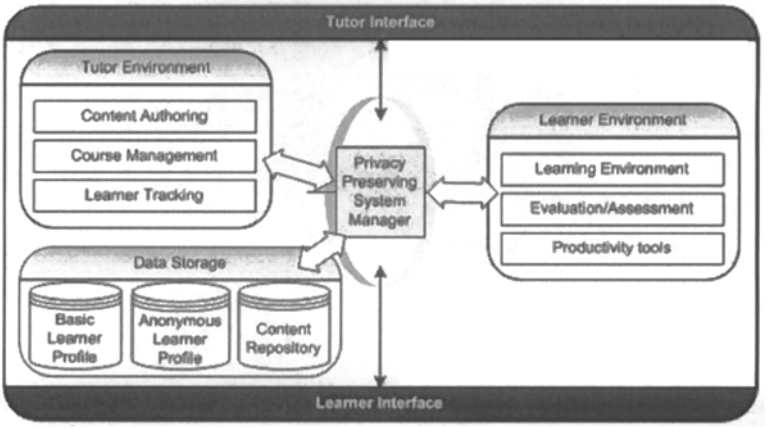

Fig. 2. Privacy-Preserving E-learning architecture

4.3 Soft Privacy learning process

Currently, we only consider the learning process for soft privacy with strong tracking. However, we leave the generalization of this process for the other levels of privacy and tracking for future work. In the context of soft privacy, the blind digital certificate (BDC) consists of the learner's public key and the encrypted form of his identity and demographic data. The learning process is as follows:

SoftPrivacyLearningProcess(Learner *U*)

1. *U* obtains a BDC from the CA
2. *U* registers in the E-learning system:
   a. Presents the BDC.
   b. Provides his learning profile information such as his goals, preferences and qualifications.
3. Repeat until the learner reaches his goals:
   a. *U* selects the courses he wishes to take.
      i. His learning profile proves that he has the necessary qualifications and skills.
      ii. His course history proves that he has the necessary prerequisites.
   b. *U* completes the course's learning activities.
      i. The E-learning system takes into account *U*'s learning profile during the learning process.
      ii. The E-learning system records the activities, and updates the learning profile accordingly.
4. Upon request
   a. *U* obtains an anonymous transcript.
   b. *U* obtains his anonymous degree.
   c. *U* updates his goals.

4.4 Satisfying the Learning Objectives

The learning process is usually defined by the learning objectives. These objectives are either professional or academic. Suppose, for instance, that the learner must prove to his manager that he completed the training course, without revealing his personal data to the E-learning system. In this case, the learning process is as follows: The learner goes through the *SoftPrivacyLearningProcess*(), and performs the required activities. Thus, at this stage, the E-learning system is in possession of the learner's BDC, and the learner's activities are evaluated by the E-learning system without compromising his privacy since his identity is unknown to the system. In particular, the E-learning system grades the learner's assignments and exams and uses its private key to sign the learner's BDC together with his results. This signed data is sent to the learner, who forwards it to his manager, along with the E-learning system's digital certificate. The manager verifies the authenticity of the E-learning system's digital certificate. He then verifies the validity of the signature on the learner's data. This last verification confirms that learning was achieved by an individual identified by the BDC. If the learner used his real identity, *rid*, when creating the BDC, then the manager can easily verify this identity by computing $E_{pk}(rid)$ and comparing this value with $E_{pk}(id)$ contained in the BDC. During the creation of the BDC, to avoid a guessing attack, the learner chooses *id* such that $id = (rid,randVal)$, where *randVal* is a random value that the learner reveals only to his manager for the BDC validation. If the learner did not use his real identity, then he provides the manager with the identifier *id*, which is encrypted as $E_{pk}(id)$ in his BDC, and the manager can then verify the authenticity of this identifier.

In addition, if the learner wishes to pursue further training activities within the same E-learning system, then he only needs to select the level of privacy/tracking. For instance, if the learner doesn't want the E-learning system to link his previous activities to new ones, he could ask for an anonymous transcript and use it to create a new account to follow new learning activities. On the other hand, if the learner decides to pursue learning activities in another E-learning system, then the current E-learning system delivers an anonymous transcript, and/or an anonymous degree to the learner. Based on these anonymous documents, the learner can prove to the new E-learning system that he possesses the required credentials and qualifications to continue his learning activities.

5 Implementation and Validation

In order to validate our approach, we implemented an E-testing prototype that supports Soft Privacy and Strong Tracking. Moreover, to keep things simple, our E-testing system only contains an Auto-Evaluation tool. At registration, the new learner is introduced to the concepts of No Privacy and Soft Privacy, and is asked to select one of them. If the new learner selects No Privacy, his Identification and Demographic data is collected and stored in the system. One the other hand, if the

learner opts for Soft Privacy, the registration will proceed as highlighted by steps 1 and 2 in the SoftPrivacyLearningProcess (Section 4.3). To implement the BDC we simulate the CA and provide the learner with a Java applet that generates the public/private key pair and encrypts his identification and demographic data at client side. Now that the learner has access to the E-testing system, he must perform at least one Auto-Evaluation test on the Java programming language. When the learner decides that he performed enough tests, he is represented with an account of the activities he performed in the E-testing system along with any information the system has on him. If the learner in question had opted for No Privacy, the system asks if he would like to switch to Soft Privacy. The purpose of this question is to determine the percentage of learners who took the issue of privacy lightly and then changed their minds after realizing the kind of information gathered about them. If the learner decides to protect his personal data, he goes through the process highlighted earlier. At this stage, learners are invited to answer a short questionnaire. First, all four privacy levels (Section 3.2) are presented and explained to the learners who are requested to choose their preferred privacy level (Fig. 3).

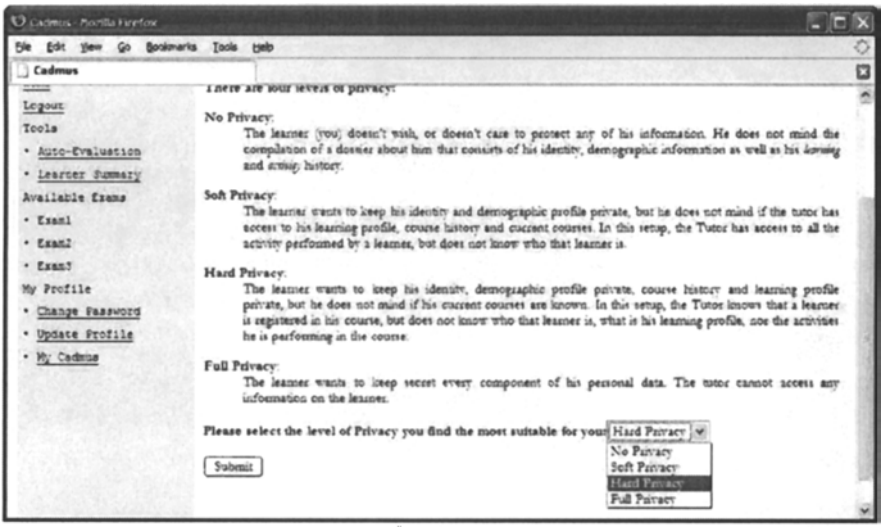

Fig. 3. Privacy selection

This question determined the interest of learners in each privacy level: none of the learners selected No Privacy and most selected Soft Privacy (Fig. 4). Afterwards, learners are introduced to the four tracking levels (Section 3.2) and they are requested to choose their preferred tracking level. Finally, learners are requested to give an overall evaluation of the system with regards to Privacy.

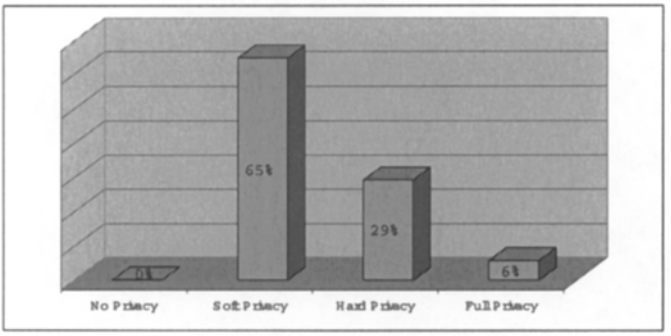

Fig. 4. Privacy Preference distribution

In summary, there were a total of 34 learners who tested the system. Among those, 82% of the learners selected Soft Privacy at registration time; 12% of the learners selected No Privacy at registration but changed their mind after seeing the kind of information the system gathers about them. Moreover, 35% of the learners who originally selected Soft Privacy selected a higher level of privacy when the four levels were introduced. Tracking preferences were almost evenly distributed across the four levels.

6 Discussion

In Section 3.2, we introduced a framework to solve privacy issues in the context of E-learning. However, we only focused on the case of soft privacy and strong tracking for the prototype implementation purpose. The hard and full privacy options are more challenging and require more cryptographic tools. For instance, in the case of full privacy, not only has the E-learning system no information about the courses currently taken by the learner, but the system must also evaluates the learner's activities for these unknown courses! Nonetheless, this can be achieved by performing the computations with encrypted functions (CEF) [13]. However, we leave the adaptation of the CEF technique, as well as a prototype for a more complete privacy-preserving E-learning system for future work.

In addition, in this work we provided tools and methods to protect learner privacy within an E-learning system. Admittedly, there are other aspects to consider: since most E-learning systems are web-based, a learner could be easily tracked through his IP address, thus violating his privacy. However, this issue can be addressed by using well-known anonymous Web surfing systems. Anonymous surfing implies that the learner can send information to and receive from the E-learning system without being traced. If the learner is in an Internet café, the learning activities can be performed somewhat anonymously. If the learner trusts a third party (such as an identity proxy [14-15]), then he can perform the learning activities from his home or office. If no single party can be trusted, Chaum's mix-nets [16] can be

used to send untraceable requests to the E-learning system, in which case an untraceable return address can serve to deliver the learning activity contents. In more general context, there is a need to address the privacy issues related to tracking as stated in Section 3.1. However, we also leave this for future work.

7 Conclusion and future work

Today, E-learning has become a standard, and there exist several virtual universities that offer online programs. Nonetheless, learner privacy is barely considered within E-learning systems, at a time when people have growing concerns about their privacy. In this work, we have presented a framework to address various learner privacy concerns in an E-learning environment. Moreover, we have implemented and tested an E-testing system to offer Soft Privacy along with Strong Tracking. Our privacy preservation approach is more robust than approaches in existing E-learning systems since, in our case, the learner alone can unlock his private data. Moreover, preliminary testing results are encouraging where 94% of the learners selected to protect their privacy, and 2 learners out of 3 who first selected No Privacy, changed their mind after seeing a sample of the data collected about them. Nonetheless, there is still work to be done in this field. Hard and Full Privacy still require implementation, while keeping in mind that E-learning systems need to gather information about the learner in order to provide a better learning experience.

Acknowledgements

We are most grateful to the three anonymous referees for suggestions that allowed us to greatly improve this paper. This gratitude covers also Professor Gilles Brassard for his valued guidance and feedback.

References

1. Arroyo, I., and Park Woolf, B.: "Inferring Learning and Attitudes from a Bayesian Network of Log File Data". *International Conference on Artificial Intelligence in Education (AIED 2005)*, pp 33–40, Amsterdam, 2005.

2. Westin, A.: *Privacy and Freedom* (Atheneum, New York, 1967).

3. Aïmeur, E., Brassard, G., Fernandez, J.M., and Mani Onana, F. S.: "Privacy-Preserving demographic filtering". *The 21st Annual ACM Symposium on Applied Computing*, pp 872–878, Dijon, 2006.

4. Mani Onana, F. S.: "Vie privée en commerce électronique". Ph.D. Thesis, Département d'informatique et de recherche opérationnelle, Université de Montréal, Mai 2006.

5. Raitman, R., Ngo, L., Augar, N., and Zhou, W.: "Security in the online e-learning environment". *IEEE International Conference on Advanced Learning Technologies (ICALT 2005)*, 5(8), pp 702–706, 2005.

6. Franz, E., Wahrig, H., Boettcher, A., and Borcea-Pfitzmann, K.: "Access Control in a Privacy-Aware eLearning Environment". *International Conference on Availability, Reliability and Security (ARES 2006)*, pp 879–886, Vienna, 2006.

7. Yee, G., and Korba, L.: "The Negotiation of Privacy Policies in Distance Education". *Information Resources Management Association International Conference (IRMA 2003)*, Philadelphia, 2003.

8. Chaum, D.: "Security without identification: Transaction systems to make Big Brother obsolete". *Communications of the ACM*, 28(10), pp 1030–1044, 1985.

9. Chaum, D., and Evertse, J.: A Secure and Privacy-protecting Protocol for Transmitting Personal Information Between Organizations. In Andrew M. Odlyzko, editor, *Advances in Cryptology – CRYPTO '86*, volume 263 of *Lecture Notes in Computer Science*, pp 118–167 (Springer, Berlin, 1987).

10. Brands, S.: *Rethinking Public Key Infrastructure and Digital Certificates – Building in Privacy*. (MIT Press, Cambridge, 2000).

11 Camenish, J., and Lysyanskaya, A.: An Efficient System for None-transferable Anonymous Credentials with Optional Anonymity Revocation. In Birgit Pfitzmann, editor, *Advances in Cryptology – EUROCRYPT 2001*, volume 2045 of *Lecture Notes in Computer Science*, pp 93–118 (Springer, Berlin, 2001).

12 Lysyanskaya, A., Rivest, R. L., Sahai, A., and Wolf S.: In Howard Heys and Carlisle Adams, editors, *Selected Areas in Cryptography*, volume 1758 of *Lecture Notes in Computer Science*, pp 184–199 (Springer, Berlin, 1999).

13. Sander, T., and Tschudin, C.: "Towards mobile cryptography". *Proceedings of the IEEE Symposium on Security and Privacy*, IEEE Computer Society Press, pp 215–224, Oakland, 1998.

14. Boyan, J.: "The Anonymizer: Protecting user privacy on theWeb". *Computer-Mediated Communication Magazine*, 4(9), 1997.

15. Gabber, E., Gibbons, P.B., Kristol, D.M., Matias Y. and Mayer A.J.: "Consistent, yet anonymous, web access with LPWA". *Communications of the ACM*, 42(2), pp. 42–47, 1999.

16. Chaum, D.: "Untraceable electronic mail, return addresses, and digital pseudonyms". *Communications of the ACM*, 24(2), pp. 84–90, 1981.