Privacy Enhanced Personalization in E-learning

Mohd M. Anwar, Jim Greer, and Christopher A. Brooks

Abstract—**It is possible to make web-based learning engaging, appealing and personalized, and yet similar to face-to-face classroom experiences by providing personal spaces to learners and teachers. Yet due to a lack of proper privacy policies and technical frameworks to implement policies, seemingly innocent data transactions can carry potential risks to privacy in e-learning environments. In this paper, we investigate privacy issues in e-learning and make recommendations for building e-learning environments that enhance privacy but allow for other features such as content personalization and collaboration with peers. We have implemented some of these recommendations, and are in the process of implementing others in iHelp, an e-learning environment that supports both learners and instructors throughout the learning process.**

Index Terms—**Anonymity, pseudonymity, identity, privacy, personalization, e-learning.**

I. Introduction

THE e-learning community is growing at a rapid pace and so are e-learners' privacy concerns. Considerable amounts of data about e-learners are being collected to provide personalized learning experiences. The collected data contain personal and sensitive information such as test scores, learning preferences, learning progress, questions asked in forums, conversations in chat rooms, and counseling sessions. As a result there are natural concerns over privacy. It is desirable to offer sufficient privacy to ensure that e-learners have autonomy in their activities and personal spaces in this relatively public educational environment.

Demchak and Fenstermacher have noted that privacy is directly related to the knowledge of identity [1]. We view identity as a dataset (e.g. name, biometric data element, behavioral pattern, etc.) that is used to model and thereby recognize an entity as distinct from others. An entity may be represented by many identity models including their own "true" identity. Naturally, some models are partial, revealing some but not all information about the entity. Some models may be incorrect – representing false information about the entity. Sometimes, a person may want to publish their own personal identity model, and sometimes they may want to keep it concealed. E-learning systems are different than many other online communities in that learners typically have more trust in the system (e.g. they are willing to part with private information readily, as they believe it will be used in evaluation), and have an extended working relationship with the system (e.g. they may work with the same forum system for many years as they progress through a program).

While personalization has become very popular in today's adaptive e-learning systems, we feel that the learner's privacy and identity management issues have largely been ignored. Kobsa and Schreck have described the risks to privacy posed by personalization [2]. The aim of this paper is to investigate privacy issues specifically in e-learning and offer solutions to address them. Any such solution must allow personalized learning, while preserving an appropriate level of privacy to the learners and teachers.

In this paper, we investigate why the privacy issues in e-learning are different from those in the traditional classroom. Extrapolating from our findings, we defined the key characteristics of an e-learning environment that provides appropriate levels of privacy, facilitates trust, and offers personalization. We have implemented many of our recommended characteristics of such an e-learning environment in iHelp[1], an e-learning environment in use at the Department of Computer Science in the University of Saskatchewan. Finally, we analyze the privacy features of iHelp to conclude that these features are helping provide privacy while allowing for advanced features (e.g. personalization, collaboration) of learning environments.

II. The Idea

Though we all intuitively understand privacy, it is a difficult thing for people to formalize and talk clearly about. In order to design a privacy preserving personalized e-learning environment, we look to deepened our understanding of the individual and social practices that constitute privacy by looking to other disciplines, in particular psychology [3], sociology [4], law [5], and computer science [6].

In psychology, privacy is defined in terms of solitude; in law, privacy is more about a control over what someone does; in sociology, privacy is about being free to behave without the risk of being observed, and in computer science, there are many definitions from the perspectives of many areas from access control to data integrity to identity management. To provide an unambiguous description of how privacy is affected

Manuscript received April 9, 2006. This work was supported in part by Science and Engineering Canada through the LORNET project.

M. M. Anwar is a PhD student in the ARIES Laboratory, Department of Computer Science, University of Saskatchewan, Saskatoon, SK, Canada (e-mail: mohd.anwar@usask.ca).

J. Greer is a professor in the Department of Computer Science, University of Saskatchewan, and director of the ARIES Laboratory.

C. A. Brooks is a research assistant in the ARIES Laboratory, Department of Computer Science, University of Saskatchewan.

[1] See http://ihelp.usask.ca for more information.

in an e-learning environment, we choose to define privacy in terms of identity.

An identity is a dataset that holds information like attributes (e.g. name, student id number), traits (e.g. interaction patterns, biometric information), and preferences (e.g. preferred meal type). An individual holds multiple partial identities in different contexts. For example, a graduate student holds multiple partial identities based on the role they play: a student, a tutor, an instructor or a marker. In the context of being in a teaching role, one's student id number may be extraneous information whereas in the context of enrolling in a class, employee id may be irrelevant. Therefore, we can say that a learner's or a teacher's privacy is their capacity to control the conditions under which their identity information will be shared.

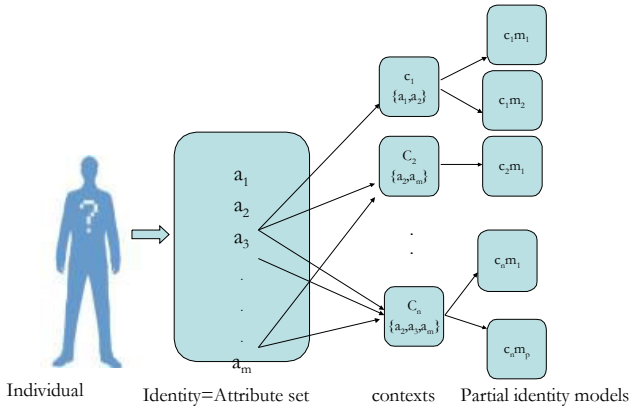

Fig. 1. Context-sensitive identity model.

Following from the notion that privacy is about protecting identity information, identity management appears to be a natural solution to privacy. Many researchers have proposed identity management schemes. Based on the amount of personal information disclosure, there are 3 approaches to identity management: anonymity (where actions may occur and be observed but no identity information is disclosed), pseudonymity (where selective identity information is disclosed by associating a pseudonym with actions that occur over time), or open identity (no restrictions to disclose identity information until some threat is observed). Although anonymity may ensure absolute privacy, it restricts personalization because no longitudinal behaviour record can be associated with an anonymous learner. Since we don't live in an ideal world, full identity is a naïve scheme, and it totally disregards privacy.

The assumptions about privacy in a traditional classroom do not apply to e-learning. A traditional classroom represents a closely knit group where everyone gets to see and know each other on a regular basis. Yet some information is protected including precise grades or confidential conversations. In contrast, e-learners and their instructors hardly get a chance to see or know each other – they are strangers whose interactions are only limited to certain selected written communications (synchronous or asynchronous). Any private information is prone to misuse if it is shared with a stranger. In a traditional classroom, physical presence works as the guarantor of authenticity whereas in e-learning, a learner needs to worry about the authenticity of their peers or instructors.

In a face-to-face communication, one can look in the eye of the interlocutor and search for tacit signs of truthfulness or falsehood [7]. The interactions in e-learning are generally devoid visual and verbal cues like facial expressions, body language, intonations, etc. In a face-to-face classroom, the instructors provide some degree of personalization by observing the visual cues of learners (e.g. a blank look of learner prompts instructor to explain a concept again). It is hard to contextualize content in an e-learning setting. In a regular classroom, teachers might say provocative things for pedagogical purposes (e.g. to incite a discussion). On the other hand, due to the lack of contextual cues in an e-learning environment, such a tactic could be taken out of context and used to misrepresent the instructor. Therefore, many seemingly non-privacy issues of regular classroom pose risks to privacy in e-learning.

We propose the following recommendations in building an e-learning environment that provide a holistic approach to preserving and protecting privacy:

- **Allow pseudonymity:** A user should be able to pick a name or let the system generate a name under which they want to be known to others. Allowing for multiple pseudonyms to be used in different contexts is also useful.
- **Allow anonymity when possible:** A user should be able to perform low risk activities (e.g. asking a question in a class discussion forum) anonymously. Anonymity can also be achieved through randomly generating a pseudonym for every transaction. Since the users tend to ignore anonymous postings, the use of transaction-based pseudonym makes the interaction more attractive.
- **Facilitate information sharing based on trust:** A user should be able to share personal information (which could include aggregate group information) with other trusted users. The system should help a user evaluate the trustworthiness of other actors and help them make informed choices about sharing personal information. Also, the system should allow its users to share some degree of social commitment. When a user A shares personal information with a user B, it is a social commitment for the user B to reciprocate in some form.
- **Allow attaching contextual cues with information:** An actor plays different roles in different contexts (e.g. socialization, collaboration with fellow learners, one-to-one communication with the instructors, etc.). The system should support context separation so that information from different contexts are not fused together to gain the knowledge of identity. Access to information needs to be controlled based on the role an actor plays in a given context.
- **Allow attaching verbal and non-verbal cues with information:** Since verbal and certain non-verbal

cues reveal the intention of an interlocutor, the system should provide a way to attach cues to information to avoid misunderstanding and grow trust among the participants of the e-learning environment. We believe that privacy is not at risk when information is shared within a trusted community. Verbal cues can be supported by providing readily available easy to attach tags (e.g. jokingly, hypothetically, sincerely, etc.), as well as emoticons (small pictorial representations of the emotional state of the user).

- **Detect and purge unnecessary personal information:** When an actor immerses themselves into e-learning activities they cannot always judge the private nature of information they share with other users. Monitoring each piece of information before its delivery is disruptive to the spontaneity of activities among the users. The system should take the job of judging the nature of information and warn the users about any accidental privacy slip.
- **Allow information to expire:** We view users as the rightful owner of their personal information even when the information is shared or observed by someone else. A user may share some of their information to somebody for a period of time. However, they should be able to destroy the information after that time is over. Sometimes, the outdated information even misrepresents a user. We suggest attaching "time to live" tag for each piece of information and making it inaccessible when the time expires. Identity Based Encryption (IBE) [8] could be used where the pseudonym of the owner of information will be used to encrypt information when the information expires. In this way, the owner of information will have the opportunity to reincarnate information when need be.
- **Promote privacy awareness:** Some users may not readily understand the need for privacy. We feel that the system should educate its users about privacy so that the users become aware of the risk of identity disclosure and learn to respect others' privacy preferences. The system could play the role of a privacy coach by providing privacy tips, presenting privacy policy of its own, and asking users for their privacy preferences
- **Punish bad actors:** To help improve responsible behaviors, users should be rewarded for maintaining a long term record of good behaviors. At the same time, the users need to be flagged with sanctions for their bad behaviors. The system could employ a reputation system that would recognize the good users with a higher reputation score. Many different levels of sanction could be applied to bad actors from restricting anonymity to revoking the privilege of participation in certain activities.

III. PRIVACY IN IHELP

We have already implemented some of our recommendations in iHelp, an e-learning environment in use at the Computer Science department at the University of Saskatchewan. We are in the process of implementing the rest of the recommendations and are proceeding with various evaluation studies. The iHelp system supports various kinds of academic roles that the users play in an academic setting (students, markers, tutorial assistants, instructors, guests, etc.) and the permissions and needs appropriate to their roles (e.g. course authoring for instructors, discussion forum for general questions or comments). Although iHelp gathers learners' information to support research on personalization (e.g. learner modeling, content improvement), the privacy features implemented in the system help it preserve privacy of its users. iHelp offers the following privacy features:

- **Pseudonymity:** An actor is given an option to select alternate pseudonyms for themselves. An actor may choose as many different pseudonyms as they want but all pseudonyms selected must be unique (distinct from one another). The system provides a default pseudonym for the users who do not choose one. When posting in the iHelp Discussion forum or chatting in a chat room under a pseudonym, an actor's true identity is not known by other actors unless they disclose their true identity.
- **Anonymity:** To welcome and engage shy learners, instructors may enable the option of anonymous posting. For anonymous postings, other users will simply see the name "anonymous" for the posters. Since the users are allowed to pick as many pseudonyms as they want, an actor may choose a new pseudonym for each transaction and enjoy virtual anonymity under the disguise of a participant with a name.
- **Context Separation:** The iHelp system facilitates context separation by providing context specific interaction channels. For example, the iHelp discussion category under the heading of CMPT_350–Assignment_1 would be open only to students in CMPT 350 as well as the instructor, teaching assistants, and other potential helpers. The iHelp system provides separate channels for asynchronous and synchronous communication corresponding to separate learning contexts. In addition, private channels for personal chat or counseling discussions between instructor and learners are available.
- **Facilitation of Trust:** As the relationship of trust grows between the actors in some context, the system does not restrict users from revealing or sharing personal information. As learners interact with one another, familiar pseudonyms emerge and attribution of personalities to pseudonyms quickly develops. Actors who are apparently helpful or knowledgeable develop a reputation but their privacy is still

protected by their pseudonym until they wish to selectively expose more identity information to those they choose to trust. In circumstances where researchers wish to relate an actor's real identity with their usage records, the researchers need to obtain informed consent from the respective users stating clearly the purpose of their study.

- **Detection and removal of unnecessary personal information:** In general, the system provides aggregate information to instructors and researchers stripping off any personal information. For example, in courses where iHelp usage counts for class participation, instructors receives summaries of a user's activities under all pseudonyms including anonymous. These summaries may include number of postings made or read, time spent, chat activity, etc. To promote our research in personalization, information are released to researchers under the strict supervision of the University. Upon ethics board approval, researchers have access to usage logs without an access to individuals' names or any other identifying information known by the system.
- **Promoting privacy awareness:** A clearly stated privacy policy allows users to make an informed choice regarding what piece of information to share with whom. At present, the system presents its privacy policy to its users.
- **Punishment for bad actors:** To keep a provision of sanction for bad behaviors, the system allows the administrators to trace the true identity of a pseudonymous or anonymous user, but only if need be. The users who misuse their disguises of pseudonymity or anonymity to commit bad actions (e.g. being abusive to others) or violate the computer use policy of our institution may lose their right to carry out activities pseudonymously or anonymously. Gross violations may result in other disciplinary actions.

IV. CONCLUSION AND FUTURE WORK

In this paper, we explored several privacy issues in e-learning. We described how any why privacy concerns are more prevalent in e-learning than in a traditional classroom. We feel that the broad acceptance and adoption of e-learning amplifies the issues we have made. We have made some recommendations in building an e-learning environment that would preserve privacy but support community building and personalization. The ability to collaborate with one another is a demand from instructors looking to implement social development theory in the classroom, and is key in making e-learning scale to large groups of users. We have reviewed our implementation of some of these recommendations in iHelp, and are now and conclude that it provides a reasonable degree of privacy protection for learners, facilitates trust, and allows personalization.

We are in the process of implementing features that support learner awareness and allow attaching verbal or non-verbal cues to information. We feel the need to build a reputation system to facilitate trust more effectively. This would involve the function of querying the reputation of a particular pseudonymous actor. We think that such a system could help users assess the trustworthiness of an actor by analyzing the quantity and quality of participation in a given context (e.g. how many times a pseudonymous actor had contributed in the discussion forum and how significant was the contribution). When the system can help users successfully identify potential good helpers or collaborators, they can work to build a relationship of trust.

Privacy protection in reputation transfer requires that the transfer must occur without letting anyone observe such transfer. We have developed a model by which this can be done with the aid of a trusted guarantor [9].

We also believe that information expiration minimizes the risk to privacy loss. Developing a model to enforce the mandatory forgetting of information seems to be very difficult. However, we plan to implement a protocol for information expiration that can be implemented within our systems using time-to-live tag for each piece of information. After the time-to-live has expired, all identity information is removed from the information.

The context of use of privacy information is an important factor in making users comfortable with sharing their various attributes. We are investigating how we can formalize context using purpose-based models of learning interactions, where a specific learning purpose (e.g. to evaluate a student, to provide help to a student, or to provide awareness of activities to another student) is mapped directly to the attributes that are required to support it (e.g. the students marks, the students learning style, or the students online activity). Integrating this into iHelp in an unobtrusive yet customizable manner is an important goal.

Anecdotal observations suggest that users use pseudonyms for a variety of reasons. We have observed that students use pseudonyms to both solicit faster response times and to imply a high level of expertise. Instructors use pseudonyms to encourage discussion, and solicit responses by posing as other students in the course. We are interested in quantifying the effects and reasons that students have for using pseudonym through both qualitative and quantitative studies.

We see privacy in e-learning as an interesting microcosm of the broader issues of privacy in online communities, and believe that the issues described here are also relevant in the online entertainment (e.g. massively multiplayer games) and business domains. The dangers associated with identity theft are not so dominant in this environment as in, say, an e-business environment, where identity theft can dwarf all other privacy concerns. Yet the issues important in privacy protection in e-learning are relevant in may other domains.

REFERENCES

[1] Demchak, C.C. and Fensternmacher, K.D. Balancing security and privacy in the information and terrorism age: distinguishing behavior

from identity institutionally and technologically. The Forum, vol. 2, pp. Article 6, 2004.

[2] Kobsa A. and Schreck J. Privacy through pseudonymity in user-adaptive systems. ACM Transactions of Internet Technology, vol. 3, pp. 149-183, 2003.

[3] Brierley-Newell, P. 1998. A cross-cultural comparison of privacy definitions and functions: A systems approach. Journal of Environment. Psychology. 18, 357–371.

[4] Altman, I., and Chemers, M. 1980. *Culture and Environment*. Wadsworth Publishing Company, Stamford, CT.

[5] Gavison, R. 1984, Privacy and the limits of law. In F. Schoeman ed. Philosophical Dimensions of Privacy: An Anthology, Cambridge University Press, New York, NY.

[6] Adams, A. 2000. Multimedia Information Changes the Whole Privacy Ballgame. In *Proceedings of Computers, Freedom, and Privacy* (CFP 2000, Toronto), ACM Press, New York, NY, 25–32.

[7] Feenberg, A. (1989). The Written World: On the Theory and Practice of Computer Conferencing. In R. Mason & A. Kaye (Eds.), Mindweave: Communication, computers and distance education (pp. 22-39). Oxford: Pergamon Press.

[8] Boneh D. and Franklin M., "Identity Based Encryption in Weil Pairing" in Proc Crypto (2001), LNCS Vol. 2139, Springer, pp. 213-229, 2001.

[9] Anwar M., "Privacy through Identity Management" in Graduate Symposium (2006). Department of Computer Science, University of Saskatchewan.