Using Students' Tracking Data in E-learning: Are We Always Aware of Security and Privacy Concerns?

Madeth May
University of Lyon
LIESP Laboratory, INSA Lyon
21 avenue Jean Capelle, Villeurbanne F-69621, France
madeth.may@insa-lyon.fr

Sébastien George
University of Lyon
LIESP Laboratory, INSA Lyon
21 avenue Jean Capelle, Villeurbanne F-69621, France
sebastien.george@insa-lyon.fr

***Abstract*—This paper presents a study on security and privacy concerns in E-learning. The study has been conducted along with our research effort that focuses on tracking students' activities on Computer-Mediated Communication tools (e.g. discussion forum, blog, wiki, etc.). It aims to express our attention on technical and ethical aspects of using tracking approach in the learning process. While the study covers an analysis of some existing research data of security in E-learning and user privacy protection provisions, it helps us gain a broader perspective of utilizing the tracking approach in our research. The major contribution of this paper is that it raises an awareness of the relevant issues, which are often neglected in the research efforts that implicate user tracking and personal data usage for instructional purposes.**

Keywords-tracking system; tracking data; computer-mediated communication; security and privacy concern

I. INTRODUCTION

The research in E-learning is no longer in its infancy, but developing very rapidly in accordance with the tremendous technological progress being made [1]. As we progress, we witness a big change of research interests in E-learning toward the improvement of technologies that better support user participation and interactivity [2]. One of the noticeable trends is on the integration of user tracking process in learning environments. In fact, using tracking systems to observe the learning process has been seen to be a reliable support to the participants, particularly in distance learning situations. For instance, by tracking students through learning environment, the tutors can keep themselves informed of the activities being undertaken and the resources being consumed by the students. It is because students' tracking data are significant sources of information that reveal both the students' activities and their outputs [3].

Using tracking system in learning environments has been steadily increasing. One of the reasons for this phenomenon is the willingness of the researchers, pedagogical teams and other practitioners to make distance learning a high quality education. Indeed, the concept of using tracking system is recognized as a contributing factor to the high quality education in terms of teaching enhancement and learning guidance. As found in [4], a review of a variety of systems that make use of learning tracking data to assist the learners in mirroring their activities and to guide them throughout the learning process. Further evidence can be found in the research works of [5-9].

Nowadays, we are confronted with a new situation. Existing technologies used in learning environments have increased security and privacy problems, which leads to a situation where security and privacy protection are becoming essential for the users. The study we present in this paper is not meant to address new research challenges, but to assist researchers, teachers and students to acquire a better understanding of security and privacy issues in E-learning. Its main objective is to raise an awareness of these issues, which are often neglected in the research efforts that implicate student tracking and the use of student's personal data. The major contribution of this study relies on the analysis of a number of existing studies, which help one gain a broader perspective on using tracking approach for the instructional purposes.

This paper is structured as follows. The second section gives an overview of our research work that focuses on an explicit tracking approach to efficiently observe the students' activities on Computer-Mediated Communication (CMC) tools. A brief discussion on why we suggest tracking approach for CMC tools and the related technical issues are presented in the same section. The third section addresses the importance of understanding security and privacy issues in E-learning. The fourth section discusses a solution to the studied issues as well as its limitation and compromise.

II. RESEARCH CONTEXT

A. Why Tracking Approach?

Users' communication activities are made on CMC tools and can be called as CMC activity in short. In distance learning, making CMC activity is not only to increase interaction among the participants, but also to compensate the lack of face-to-face interaction. According to [10], CMC tool is recognized as an essential element to online learning situation and is strongly recommended for the participants.

Even though existing research works proved that using CMC tools enhances online teaching and learning, there are still issues that we should recognize. If we take a closer look at the use of CMC tool in distance learning, CMC tool alone does not always enable the participants to fully control their activities the way they do in a traditional face-to-face learning situation. As a matter of fact, the interactions

978-1-4244-9587-0/11/$26.00 ©

between the participants are not person to person, but computer-mediated and online, which makes it difficult, for example, for the teachers to supervise the students' activities. As for the students, they could easily encounter difficulties in self-monitoring if CMC tool was the only support they had for conducting their learning activities. This is due to the fact that CMC tools, from a technological standpoint, were not originally built to assist the teachers in monitoring the students. Besides, CMC tools do not provide technical assistance to the students to gain an insight of their activities and those of others. Another issue found in online learning is that the students often receive supports that strongly rely upon their teacher's commitment and are usually constrained by other factors related to distance and time. Meanwhile, with the current support of CMC tools that are often limited to communication means, the participants are compelled to neglect some fundamental facets of online learning, such as self-monitoring and self-evaluation.

Having studied these issues, we addressed the importance of tracking CMC in learning situations for the benefits of tracking data to online tutoring and learning enhancements. An explicit tracking approach has been proposed for the implementation of tracking systems for a great variety of CMC tools. It focuses on a tracking mechanism capable of observing different types of user action and interaction on CMC tools. Later, we continue our research by focusing on exploiting the collected data to support the participants in terms of gaining awareness and making assessment of their learning activities, outcomes, effectiveness, etc.

B. Technical Issues in Tracking CMC Activity

In order to efficiently track users' communication activities on CMC tools, the tracking system must closely follow the activities taking place. However, in the existing tracking methods, most systems were designed to observe the users activity only on the server side (e.g. where the communication platform is hosted). The user interaction on the client side (e.g. user Web browser) is often ignored. In this method, the granularity of the tracking data should be rather large and may not be accurate enough to reflect the complete activities of users on CMC tool. While tracking data collected from the client side are either left behind or incomplete, they represent the behavioral aspect and/or the process of user interaction during an activity. On top of that, the collected tracking data are used to reflect the actual users' activities. Hence, they should contain significant information that describes both the process and the product of the activity.

An attempt has been made to investigate the problems, "how to design a tracking systems capable of efficiently tracking users' activities on both client and server side?", and "how to make tracking data useful to both teachers and students?"

C. A Tracking Approach for CMC Tools

Users' activities on a CMC tool consist of a large part of Human and Computer Interactions, which are technically the "observable objects" traceable by the tracking system. Therefore, the proposed approach focuses on the observation of users' activities at different levels of interaction, as shown in figure 1.

(1) The Human–Computer Interaction (HCI) refers to the user's actions while using the CMC-tool interface, which occur only on the client side. If we look at an example of an activity "writing a new message" on a discussion forum, the user interactions can be: "edit" message title or message content, "move" vertical scrollbars upward or downward, "drag and drop" smilies into the message, etc. The main reason of tracking on the client side is that HCI tracking data are beneficial in identification of user's behavior while using a CMC tool to perform a communication activity. They are also compulsory in the process of rebuilding successive user's actions and events of the past activity (e.g. what did a user do to write a new message).

(2) The Human–Human Interaction Mediated by Computer (HHIMC) refers to the content of the interaction among users. With the same example of "writing a new message" on a discussion forum; all the written text as well as the attachments will be submitted to the server so that the message can be read by other users. The collected tracking data of HHIMC will be exploited along with those of HCI to make the data more descriptive and to enable an awareness of both the process of an interaction (e.g. how a user writes a new message) and its product (e.g. what the message is about).

(3) The Computer–Computer Interaction (CCI): keeping track of meaningful events means to track both the computer input and output processes while a communication takes place. The tracking data of CCI are very useful for the designers and developers who seek to improve the CMC tools; and for the researchers who are involved in development experiences. For instance, developers commonly use the CCI tracking data to debug problems related CMC tools and to strengthen the security of the communication.

(4) The Non–Computer Mediated Human Action (HA): this covers all users other actions outside the computer environment (e.g. a user makes a phone call during the learning session). In some circumstances, particularly in remote situations, it is not sufficient to track only the computer-mediated activities of the users. As yet, video and audio recorders are more practical in observing what cannot be observed by the computer-mediated tracking system. It

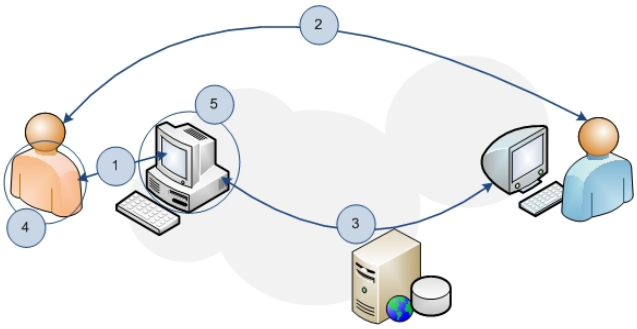

Figure 1. Different types of actions and interactions of a CMC activity.

should be noted that we are not using the audiovisual data in the current research stage.

(5) The Computer Action without Human Action (CA): there are many computer actions that occur automatically without the action of the user. Examples include a popup message indicating that a user session in a chat room will expire in 5 minutes. The tracking data of computer action usually describe what else happens besides the HCI. That is why such data are often used as supplementary information to complete the tracking data of actions and interactions presented earlier.

D. Example of Tracking Data Usage

Exploiting tracking data is a complex process. It involves many phases among which the transformation of tracking data into graphical forms, allowing users not only to easily visualize the data but also to interpret them.

Figure 2 gives an example of tracking data visualization. It illustrates different activities of two students on a discussion forum. Each radar graph displays quantitative data of (a) the discussion threads that the student started, (b) the messages quoted, posted and replied by the student, (c) the files that the student uploaded and downloaded, and (d) the student's participation level. In practice, such visualization serves multiple purposes, among which the analysis of various aspects of students' interactions on the forum. First, figure 1 supports the comparison of the

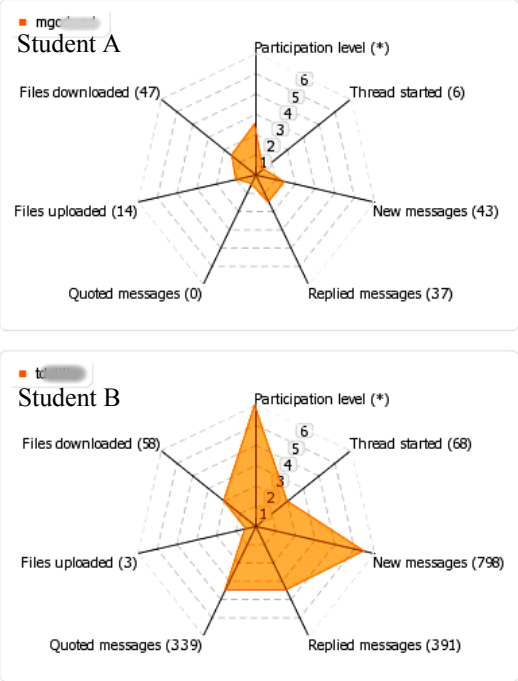

Figure 2. An example of visualizing students' tracking data.

participation level between two students and their contribution to the group discussion. More precisely, the student participation level can be determined by different percentages of (i) forum browsing, which indicate the activeness of a student on the forum, (ii) message posting and (iii) message reading activities. The contribution of a student, on the other hand, can be identified by a set of information, including new postings and documents realized by the student and shared with others on the forum (i.e. `New messages` and `Files uploaded`).

Second, figure 2 also leads to an identification of the level of social interaction of each student. For instance, the number of threads a student started could reflect the interest of the student in making discussions. Meanwhile, the number of messages a student quoted and replies could reveal how active the student was in interacting among other students. Further discussion and examples of tracking data visualization can be found in our recent published work [11].

III. A STUDY ON SECURITY AND PRIVACY ISSUES IN E-LEARNING

Understanding the security issues in learning situations helps the participants to avoid security threats as well as to improve protection of both participants and their learning environments [12]. In E-learning situations, according to [13], privacy issues concern learning technology providers, learning service and content providers, and the participants themselves. Indeed, the crucial tasks for learning service and content providers are to secure learning environment and to secure storage of learner data. As for the participants, they are mainly concerned with trust assessment of learning environments they are using, and with protection of their sensitive personal data [14].

Security and privacy levels differ in various learning environments and depend on types of learning activities being conducted by the participants. We witness that in a collaborative learning situation where interactions between participants are inevitable and their exchanges of both personal and collaborative data are intense, a strong protection of participant's privacy could only be done on a particular environment that is specifically built for such situation. As found in [15] on establishing a privacy-aware collaborative learning environment and [16] on multi-dimensional privacy protection for digital collaborations, allowing users to perform collaborative learning activities with a high-level protection of user privacy.

To have an overview of some issues of privacy and security in learning technology as well as learners and their protection provisions, we look at some research data taken from [13] and [17]. These research efforts studied different topics related to security and privacy issues in Technology Enhanced Learning. Figure 3 reflects the urgency of different protection provisions of the following issues: personal data protection, anonymous use of learning services, address and location privacy, single sign-on, seamless access to learning resources, authenticity of learning resources (LRs), digital rights management, legislation and awareness raising. Figure 4 depicts the average of the privacy issues protection provisions.

	Non relevant	Nice to have	Relevant	Urgent	Very urgent	I don't know	No answer
Protection of personal data	0%	9%	36%	31%	24%	1%	0%
Anonymous use	10%	16%	45%	23%	6%	1%	0%
Address and location privacy	5%	10%	37%	22%	22%	1%	2%
Single sign-on	3%	12%	37%	26%	16%	2%	5%
Seamless access	1%	10%	33%	33%	17%	1%	5%
Authenticity of LRs	2%	14%	25%	32%	24%	1%	2%
Digital Rights Management	15%	11%	33%	25%	16%	0%	0%
Legislation	7%	13%	52%	18%	5%	1%	3%
Awareness raising	5%	10%	26%	24%	33%	1%	2%

Figure 3. Urgency of protection measures.

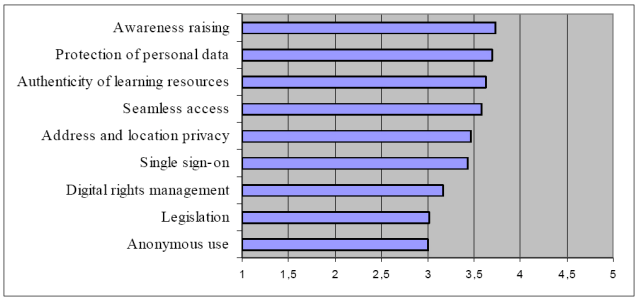

Figure 4. Average values of urgency of different protection measures.

The synthetic information in both figure 3 and 4 is computed from a questionnaire data on people's satisfaction with current security and privacy in E-learning, a view on future E-learning security and privacy, and urgency of different protection measures. A total of 147 people responded to the questionnaire, among which 66% represented universities and higher educational institutions. 67 participants are learning technology and service providers, 38 are learning content providers, and 42 are end-user organizations.

Interesting information can be retrieved from figure 3 and 4. Examples include user data protection and anonymity that are strongly relevant to privacy concern in learning environment (cf. second row of figure 3 and second horizontal bar of figure 4). Besides personal data protection, students requested to be able to control the visibility of their sensitive data such as history of their learning activities and their profiles. That is why various privacy-enhancing technologies are proposed by [18] and [19] for privacy protection at both learner side and provider side. Those technologies include identity protectors, anonymous communication systems and cryptographic mechanisms.

Regarding the tracking process of learner's activity in learning environments, [13] pointed out that 55% of end-users perceive user tracking as a big or very big threat. Interestingly, we have found similar results in the study of [20] that user tracking is not welcome even when users receive personalized content in return. Similar results were obtained for unsolicited profiling (45%) and personalization (40%) in the research work of [13].

To wrap up, this study enables us to gain an insight of the most crucial aspects regarding the security and privacy concerns in E-learning: the awareness of users when being tracked and the protection of their personal data. It also inspires us to explore a proper solution for our research.

IV. Discussion on Solution to the Security and Privacy Concerns

A. Trust is Part of the Solution

To get the better of privacy concerns is not only about using technological solutions to keep users safe from any threats, but also about "trust". According to [21], trust is a confidence in someone's competence and his or her commitment to a goal. Trust is also a crucial enabler for meaningful and mutually beneficial interactions that build and sustain learner collaboration and community [14]. As yet, privacy is a natural concern at the same time that trust is an important factor in learning environment because in practice, privacy and trust are circularly related. In reality, in a closed learning environment, where all learning services are provided internally (e.g. from a university or a trusted source) students can have higher confidence that their personal data will be treated properly. Thus, their learning tasks such as working collaboratively with other learners could be effectively conducted upon trust [22], [23]. On the other hand, in an open learning environment with unknown providers such as private or external learning service providers, privacy concerns are higher and the trust level of learners will be influenced by the level of perceived privacy offered by those providers. So finally, privacy and trust complement each other, and together they can make for a more stable learning community [24].

B. Limitation and Compromise

Regarding the discussed issues, we have two different perspectives. From a technological perspective, the solution to the security and privacy issues is still heavily reliant on technological approaches. We are convincing that better privacy protection tools are required in learning environment to manage and safeguard learning tracking data and personal information of the participants. The solutions, as pointed out [25] will involve the development and integration of Privacy Enhancing Technologies such as identify protector [26], shield privacy [27], and privacy protector [28]. Those technologies are largely used to strengthen user privacy and to secure learning environments. Thus far, we are strongly interested in exploring and exploiting those technologies in our research effort. As a matter of fact, acknowledging the ethical aspects of using students' tracking data and assuring their protection has always been our preoccupation.

From a researcher in E-learning perspective, what is important is the fact that student's personal data benefit from any type of exposure in any circumstance. Nevertheless, a compromise between tracking students and protecting their privacy is still needed. For example, allowing students to anonymously access to their learning environments for a privacy reason is feasible from a technological standpoint, but somehow limited from the fact that a learning application aims at assisting students and so they can not act in full anonymity [29]. For that reason, we suggest that learning application researchers, designers, developers and administrators should be aware of privacy requirements in their applications, from both legal point of view and as a way of ensuring students' concerns on their data protection.

V. CONCLUSIONS

This paper presents a study on security and privacy concerns in E-learning, which are often neglected in the research efforts that use tracking approach to enhance online teaching and learning experiences. While it is mainly to express our attention on privacy threats and data protection in E-learning, it aims to raise an awareness of researchers, pedagogical teams and other practitioners in terms of student tracking and personal data usage. It also discusses existing solutions along with our perceptions on the limitation and compromise when it comes to actual learning practices.

To conclude, using tracking approach in our research should not be seen as a threat to the participants for the following reasons. Firstly, we always inform users of any tracking process when they access learning platforms or use CMC tools. Secondly, only on approval of users that any tracking process can take place. Besides, there is always an acknowledgement from our part on the protection of users' personal data and their entity privacy. On top of that, users also have a full control on their tracking data and especially they have the right to make their data accessible or not by others. Last but not least, every use of the learning data we acquire in this research is strictly for educational purpose only. We are currently conducting a questionnaire on using tracking approach in collaborative learning in order to investigate the evolution of privacy concerns after all these years. We also expect that the responses from the questionnaire could serve for a study on the impact of security and privacy threats in authentic learning situations.

REFERENCES

[1] P. Scott and C. Vanoirbeek, "Technology-Enhanced Learning," Technology-Enhanced Learning, vol. 71, pp. 12-13, 2007.

[2] P. Manson, "Technology-Enhanced Learning: Supporting Learning in the 21st Century," Technology-Enhanced Learning, vol. 71, p. 3, 2007.

[3] M. May, S. George, and P. Prévôt, "A Closer Look at Tracking Human & Computer Interactions in Web-Based Communications," International Journal of Interactive Technology and Smart Education, vol. 5, no. 3, pp. 170-188, 2008.

[4] P. Jermann, A. Soller, and M. Muehlenbrock, "From Mirroring to Guiding: A Review of State of the Art Technology for Supporting Collaborative Learning," in Proceedings of the First European Conference on Computer-Supported Collaborative Learning, pp. 324-331, 2001.

[5] V. Komis, N. Avouris, and C. Fidas, "Computer-supported collaborative concept mapping: study of synchronous peer interaction," Education and Information Technologies, vol. 7, no. 2, pp. 69-188, 2002.

[6] C. Després, "Synchronous tutoring in distance learning," in Artificial Intelligence in Education, pp. 271-278, 2003.

[7] J. Hardy, M. Antonioletti, and S. Bates, "E-learner tracking: tools for discovering learner behavior," in IASTED International Conference on Web-based Education, pp. 458-463, 2004.

[8] R. Mazza and L. Botturi, "Monitoring an Online Course with the GISMO Tool: A Case Study," International Journal of of Interactive Learning Research, vol. 18, no. 1, pp. 251-265, 2007.

[9] G. Dyke, K. Lund, and J. Girardot, "Tatiana: an environment to support the CSCL analysis process," in Computer Supported Collaborative Learning, pp. 58-67, 2009.

[10] Z. Berge and M. Collins, "Computer-Mediated Communication and the Online Classroom in Distance Learning," Computer-Mediated Communication Magazine, vol. 2, no. 4, p. 6, 1995.

[11] M. May, S. George, and P. Prévôt, "TrAVis to Enhance Online Tutoring and Learning Activities: Real Time Visualization of Students Tracking Data," in IADIS International Conference on E-learning, pp. 57-64, 2010.

[12] E. Weippl, Security in E-Learning, Advances in Information Security., 16 vols. USA: Springer, 2005.

[13] T. Klobucar, M. Jenabi, A. Kaibel, and A. Karapidis, Security and Privacy Issues in Technology Enhanced Learning, ISO Press. Amsterdam: IOS Press, 2007.

[14] M. Anwar and J. Greer, "Reputation Management in Privacy-enhanced E-learning," in Proceedings of the 3rd Annual Scientific Conference of the LORNET Research Network, p. 6 pages, 2006.

[15] K. Borcea-Pfitzmann, K. Liesebach, and A. Pfitzmann, "Establishing a Privacy-Aware Collaborative eLearning Environment," in Proceedings of the EADTU Annual Conference 2005: Towards Lisbon 2010: Collaboration for Innovative Content in Lifelong Open and Flexible Learning, p. 8 pages, 2005.

[16] G. Skinner, "Multi-Dimensional Privacy Protection for Digital Collaborations," International Journal of Security, vol. 1, no. 1, pp. 22-31, 2007.

[17] M. Wolpers and G. Grohmann, "PROLEARN: Technology Enhanced Learning and Knowledge Distribution for the Corporate World," International Journal of Metadata, Semantics and Ontologies, vol. 1, no. 1, pp. 44-61, 2005.

[18] V. Senicar, B. Jerman-Blazic, and T. Klobucar, "Privacy Enhancing Technologies – approaches and development. Computer Standards & Interfaces," Computer Standards & Interfaces, vol. 25, no. 2, pp. 147-158, 2003.

[19] K. El-Khatib, L. Korba, Y. Xu, and G. Yee, "Privacy and Security in E-Learning," International Journal of Distance Education, vol. 1, no. 4, p. 16 pages, 2003.

[20] S. Fox, L. Rainie, J. Horrigan, A. Lenhart, T. Spooner, and C. Carter, Trust and privacy online: Why Americans want to rewrite the rules. Washington DC, USA: , 2000.

[21] C. Handy, "Trust and the Virtual Organization," Harvard Business Review, vol. 73, no. 3, pp. 40-50, 1995.

[22] J. Nickel and H. Schaumburg, "Electronic Privacy, Trust and Self-Disclosure in e-Recruitment," in Proceedings of ACM CHI - Computer Human Interaction, pp. 1231-1234, 2004.

[23] J. Mason and P. Lefrere, "Trust, Collaboration, and Organisational Transformation," International Journal of Training and Development, vol. 7, no. 4, pp. 259-271, 2003.

[24] J. Steel, "Interpersonal Correlates of Trust and Self-Disclosure," Psychological Reports, vol. 68, no. 1, pp. 1319-1320, 1991.

[25] R. Davison, R. Clarke, J. Smith, D. Langford, and B. Kuo, "Information Privacy in a Globally Networked Society: Implications for IS Research," Communications of the Association for Information Systems, vol. 12, no. 1, pp. 341-365, 2003.

[26] G. van Blarkom, J. Borking, and J. Olk, Handbook of Privacy and Privacy-Enhancing Technologies, Privacy Incorporated Software Agent (PISA) Consortium. The Hague, The Netherlands: College bescherming persoonsgegevens, 2003.

[27] G. Skinner and E. Chang, "A Conceptual Framework for Information Privacy and Security in Collaborative Environments," International Journal of Computer Science and Network Security, vol. 6, no. 2, pp. 166-172, 2006.

[28] D. Gritzalis, "Embedding privacy in IT applications development," Information Management and Computer Security, vol. 12, no. 1, pp. 8-26, 2004.

[29] A. Pfitzmann and M. Hansen, "Anonymity, Unlinkability, Unobservability, Pseudonymity, and Identity Management - A Consolidated Proposal for Terminology - Draft status," 2005.