Chapter 23
Preserving Learners' Privacy

Esma Aïmeur and Hicham Hage

Département d'Informatique, Université de Montréal, Pavillon André-Aisenstadt,
CP 6128 succ. Centre-Ville, Montréal QC, H3C 3J7, Canada
{aimeur,hagehich}@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. Existing E-learning standards offer some provisions for privacy and the security of E-learning systems offers some privacy protection. Privacy preserving E-learning solutions fall short and still require further development. Additionally, the advent of E-learning 2.0 introduced a whole new set of challenges with regards to privacy preservation. In this chapter we introduce E-learning systems security and privacy preserving approaches, challenges they still face, as well as the challenges brought forth by E-learning 2.0.

23.1 Introduction

When E-learning first emerged, 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. With E-learning, education is shifting from being Tutor Centered, where the tutor is the center and has access to the resources, and becoming more Learner Centered (Mccombs and Vakili 2005), where the student is the center and the focus of the learning process and has access to a multitude of resources. Although learner centered education is not a novel idea, E-learning, whether by using an LMS (Learning Management System) or an ITS (Intelligent Tutoring Systems) (Brooks et al. 2006; Woolf, 2008) are major contributors to the development and advancement of learner centered education. Indeed, one of the main drives behind E-learning is to personalize the learning experience to the individual learner. As such, in order to tailor the learning experience, E-learning systems take into consideration various factors including the learner's level of knowledge, reasoning method, preferred learning style, cultural background, even the learner's emotional state (Blanchard et al. 2009; Conati 2002; Dolog et al. 2004).

In order to provide such a level of personalization, E-learning systems collect large amounts of information about the learner, information that could be misused, and therefore violating his/her privacy. There are many reasons why learners might need to keep private different parts of their profile, and existing research

R. Nkambou et al. (Eds.): Advances in Intelligent Tutoring Systems, SCI 308, pp. 465–483.
springerlink.com

(Aïmeur et al. 2007; Anwar & Greer 2009; Hage & Aïmeur 2009a) indicates that learners have a preference for privacy in E-learning and tend to perform better. Existing E-learning standards offer some provisions for privacy and the security aspects of E-learning systems do offer some privacy protection; nonetheless it remains unsatisfactory on several levels. On the other hand, privacy preserving E-learning solutions, such as (Aïmeur et al. 2007) and (Anwar and Greer 2009) do satisfy the privacy constraint, but come at a price. Moreover, these solutions are not adequate for E-learning 2.0 and PLEs (Personal Learning Evnironment) (van Harmelen 2006). In particular, with the availability of the numerous tools which are available to the learners, tools that are external to the E-learning system and out of its control, it becomes difficult to protect the learners' information and privacy, which represents a new set of challenges. This chapter highlights the importance of security in E-learning systems, and corroborate the need for privacy. Moreover, it details some approaches to privacy preserving E-learning, their shortcomings and the challenges that still lay ahead. This chapter also provides an introduction to E-learning 2.0 and the new challenges it brings with regards to learner privacy.

The chapter is organized as follows: the next section provides an overview of security in E-learning systems and provides an overview of some common existing threats. The next section introduces privacy preserving E-learning, why privacy is important, some approaches to insure privacy and the challenges to be solved. The next section provides an introduction to E-learning 2.0 and highlights the new challenges it raises with regards to preserving learner privacy, and the last section concludes the chapter.

23.2 Security of E-Learning Systems

Security is an important aspect of E-learning. Indeed, most (if not all) of the E-learning systems and Intelligent Tutoring Systems store information about the learner, and use an underlying layer of communication between the client computer (where the learner is working) and the server (where the application is actually running). In this section we first introduce some notions about security, and then we highlight some underlying threats that need to be considered, from a security point of view.

23.2.1 Pillars of Security

Information security, (in this case the learners' information) is based on three pillars: *Confidentiality*, *Integrity*, and *Availability*. Maintaining the Confidentiality of the information involves protecting the information from unwarranted disclosure, and making sure that only the users with the proper privileges have access to that information. In other words, the user can only access the information he is permitted to. On one hand, the confidentiality of the information is considered during the transfer of the data between the client and the server. Indeed, with the availability

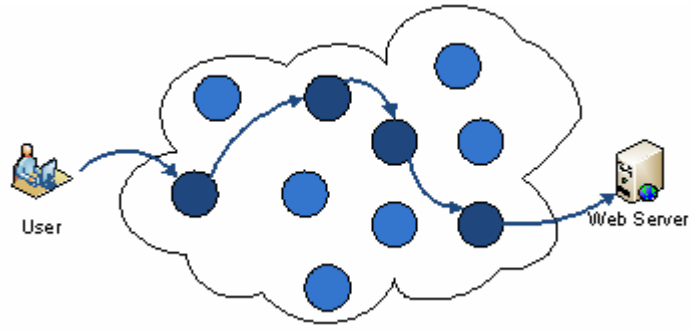

Fig. 23.1 information route from user's PC to web server

of high bandwidth and the speed of the internet connection we tend to forget that in order to reach a certain website, the connection goes through several connection points. Indeed, running a simple "*tracert*" to the website we are trying to reach displays the detailed information about the route taken by any information exchanged between the user's PC and the web server hosting that certain web page. Fig. 23.1 highlights such a route, where the circles in the cloud illustrate possible connection points.

Consequently, imagine a learner sending his login information, or even uploading his homework to the e-learning system: the data could be intercepted and used maliciously by another learner. Similarly, imagine the learner requesting his grade report for the e-learning system: that information could be viewed by an unauthorized person while being sent from the server to the learner. On the other hand, the confidentiality of the information is also considered while it is being stored within the system. Indeed, imagine that any person with access to the registrar's office of your academic institution can also access and view your academic record. Regardless whether you have a good or bad academic record, this is inacceptable. Similarly, the confidentiality of the information stored within the E-learning systems should be guarded, and only the persons with the proper access privileges might have access to that information.

The second pillar is Integrity, which enforces the validity and authenticity of the data. In other words, ensuring information integrity protects the data from any tampering or modifications from unauthorized users. To begin with, the integrity of the information is considered during the transfer of the data between the client and the server. Indeed, consider taking a learner taking an online quiz. The answers to the quiz's questions are sent through the same route described earlier. Without any integrity verification mechanisms, to insure that the data was not modified through the transmission, a malicious user can intercept the answers of the learner and modify them before they reach the e-learning system, successfully tampering with the learner's score. Additionally, the integrity of the information is also considered while it is being stored within the system. Indeed, again in this case, without the proper mechanisms to protect the data integrity, a malicious user with access to the e-learning system could tamper with the information (increasing or decreasing a test score for instance) unnoticed.

The third pillar is Availability, which relates to the availability of the e-learning system. Indeed, such systems must be available at all time, and provisions must be considered to implement and ensure this availability. One might be tempted to think: why is this crucial? Well, consider a learner without access to the system the day homework is due. Not only will the learner not have access to any necessary learning resources available through the e-learning system, but he will not be able to submit his homework. Moreover, consider a learner performing an online quiz. Even if the system was not available for only a few minutes, it is still precious time lost, not to mention the stress and the emotional pressure caused to the learner.

23.2.2 Security Threats

This section highlights *some* of the existing security threats. Actually, what we list here is the tip of the iceberg, and is intended to raise the awareness that when on the internet, we are not as safe and secure as we think we are.

23.2.2.1 SQL Injection

SQL Injection exploits security vulnerabilities at the database level of the system. Such vulnerabilities occur when the user input (data provided by the user) is not properly filtered, allowing the user input to contain executable SQL code. For instance, consider an authentication system that asks the user to provide a user name and a password, and uses the following query to validate the user's credentials:

```
SELECT * FROM user_table
WHERE      user_name = provided_user_name
AND        user_password = provided_password".
```

If the user enters valid values for the variables *provided_user_name* and *provided_password*, the query will work just fine and as expected. Nonetheless, if a malicious user provides the following user name: *"abcd OR 1=1 --"*, the *WHERE* clause of the query becomes: *"WHERE user_name = abcd OR 1=1 -- AND user_password = provided_password"*. In this case, regardless of the password provided by the malicious user, since the *"--"* is a comment in SQL, the database system will ignore anything that comes after it. Consequently, the query will always return the entire users list from the *user_table* due to the *"OR 1=1"* in the query.

Although the example portrayed here is fairly simple, malicious users using the SQL Injection attack can formulate far more complex queries and do a large amount of damage. Indeed, just to cite a couple of recent events, in August of 2009, the BBC published a story about a US citizen allegedly stealing 130 million credit card numbers using an SQL injection attack (BBC 2009b). More recently, in December of 2009, the New York Times reported on a hacker who accessed, using an SQL Injection attack, the RockYou (rockyou.com) database where he

found unencrypted login information for more than 32 million user accounts (O'Dell 2009).

The consequences of a successful SQL Injection attack on an e-learning system are numerous: the attacker could have access to the tutor's resources (upcoming exams or homework, grade books, etc.) or the learner's resource (homework, reports, learning resources, etc.).

23.2.2.2 Cross Site Request Forgery

Cross-Site Request Forgery (CSRF) is an injection type attack, where a malicious web site causes the user's browser to perform unwanted actions on a trusted site. Specifically, the malicious website would try to inject *malicious* requests to the trusted website. For example, consider a user that is logged in to his banking website to pay his bills, while at the same time browsing the malicious website. The malicious website could send a request to the banking site, asking for a money transfer to a specific account held by the attacker. Specifically, the malicious website could post an image that links to the website banking site instead, using the following link for example:

"http://mybank.com/transfer?from=account&amount=1000&to=malicious"

It is important to note that such attacks are difficult: the attacker must first gather different information about the targeted site, and the targeted user. Moreover, in order for this attack to happen, the user must be simultaneously have a valid session opened on the targeted site, and be connected to the site of origin of the attack. Nonetheless, these vulnerabilities are real and could have a devastating effect. In this report (Zeller and Felten 2008), a professor from Princeton and his graduate student report on successful CSRF attacks against several popular websites, including ING Direct (ingdirect.com), where they were able to transfer funds out of users' accounts.

A CSRF attack can be used to manipulate the E-learning system into releasing, modifying or even deleting sensitive information. For instance, a learner could manipulate the E-learning system into modifying the grade book in order to increase his own grades.

23.2.2.3 Denial of Service

A *denial-of-service* attack (DoS attack) or *distributed denial-of-service* attack (DDoS attack) is an attempt to overload a computer's resources in order to render it unable to process legitimate users' requests. It is generally conducted against web servers, saturating them with *fake* requests, making them unable to process genuine users' requests. One common method of attack involves overwhelming the target machine by saturating it with *fake* communications requests, such that it cannot respond to legitimate request, or responds so slowly as to be rendered effectively unavailable. A distributed denial of service attack (DDoS) occurs when multiple systems collaborate to flood the resources of the targeted system. Often,

DDoS attacks are conducted using *zombie* machines, computers that were compromised and are now being controlled by the attacker.

In July 2009, South Korea witnessed one of its the largest cyber attacks. DDoS attacks were used to crash the websites of dozens of government offices and banks among others (Lee 2009). Additionally, in August of 2009, Twitter and Facebook were the victims of similar attacks. While Twitter was taken offline for a while by the attacks, Facebook's service was reduced (BBC 2009a). Such attacks are quite common and usually used for extortion purposes (Messmer 2010).

Such an attack could also affect the E-learning systems in various ways: slowing down the system during an exam, or even completely crippling the E-learning system effectively disrupting any learning activity.

23.3 Privacy Preserving E-Learning

One of the main advantages of E-learning and Intelligent Tutoring Systems is their adaptability to the learner's specific needs and preferences. Nonetheless, to do so, these systems collect large amounts of information about the learner, information that could be misused, and therefore 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 (Westin 1967).

Although the security of E-learning system is imperative to preserve privacy, it is not enough. Indeed, security will protect learners' information against unwarranted access, but not against abuse from authorized access. Specifically, the insuring the *Integrity* and *Confidentiality* of the learner's information does protect the learner's data (and consequently his privacy) from unauthorised access, nonetheless, E-learning systems gather large amounts of information about the learners, information that is readily made available for the tutor, or even E-learning platform system administrator.

Specifically, *privacy* is nearly absent in current E-learning systems. Only primitive forms of privacy are offered in some platforms, for instance not allowing tutor access to certain information such as auto-evaluations performed by the learners. Nonetheless, the tutor has access to virtually all the remaining information including, but not limited to, who the students are, what parts of the course they referred to, how many times and for how long, as well as all the messages in the forums, and all the information about the quizzes and tests the learner took in his course. While learners' privacy is largely ignored within E-learning, it remains an important aspect for learners.

23.3.1 Why Privacy Preserving E-Learning

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 *personal*, or *competitive* reasons. In the **Competitive** context, the learner requires his privacy due to competitive considerations. For example, consider a prominent politician taking a course to increase his

knowledge in a certain domain of interest to the electors. 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. On the other hand, 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.

Indeed, existing research demonstrates the effect of emotions on learning (Zins et al. 2007): positive emotions improve the performance whereas negative emotions hinder the thought processes. Additionally, studies are conducted to evaluate the impact of various factors on the learner's emotional state. The purpose of these studies is to avoid situations which create negative emotions, while motivating the occurrence of situations which create positive emotions.

Fig. 23.2 Capturing the participant's most dominant emotion (Hage and Aïmeur 2009b)

In tandem, learners have conveyed a clear preference to privacy in E-learning systems (Aïmeur et al. 2007), and reported being more comfortable engaging in course related forums in privacy preserving mode (Anwar and Greer 2009). Additionally, in a recent study (Hage and Aïmeur 2009a), we investigated the impact of privacy on the leaner's emotions, and whether privacy had a positive or negative impact on learners. Specifically, in this the study, we attempt to determine, in the context of a web-based assessment, whether privacy would have a positive effect (effectively reducing stress and helping learners perform better), or a negative effect (learners would become reckless and careless about their grades). There was a total of 77 participants in the experiment which consisted of two IQ tests, one performed in a traditional none private environment, and the other in a privacy preserving environment. In order to preserve the privacy of the participants, a *random id* (*rid*) was created to and used instead of the actual participants' identifier. Consequently, the participants were informed that all their actions within the privacy preserving environment are recorded using the *rid* which cannot be linked back to them. Nonetheless, in order for us to be able to evaluate the impact of privacy, we had to *deceive* the participants and maintain an actual link between the participants profile and his *rid*.

Moreover, the participants' most dominant emotion was recorded before and after each test (Fig. 23.2) in order to determine the effect of their emotions on the score as well as their attitudes towards their performance in the tests.

In summary, on average, participants performed better in the privacy preserving test (higher score and lower average response time). Additionally, the effect of the negative emotions on the performance of the participant was lower in the privacy preserving environment. In details, the participants were separated into two groups: the first group was composed of the participants who reported a *positive* emotion prior to the test, and the second was composed of the participants who reported a *negative* emotion prior to the test. We then compared the averages of each group. The group which reported a positive emotion, on average, performed better on both tests (with and without privacy). Hence, privacy preserving E-learning is not just necessary to protect learners' information, but can also enhance the learning experience. Consequently there were some proposed approaches.

23.3.2 Existing Approaches

E-learning systems use information about a learner in order to adapt the learning activity and the interactions of the E-learning system. Such information is referred to as the learner profile or learner model. Many E-learning systems use their own internal representation of the learner model. Nonetheless, there are several standards and specifications to represent the learner model, including the IEEE LTSC Personal and Private Information draft standard (LTSC) and the IMS Learner Information Package (IMS). Although these specifications contain some attributes and means that may uphold learner privacy, the detailed specification is still missing. Moreover, the learner involvement in deciding which information is private or not is not enabled (Jerman-Blazic and Klobucar 2005), consequently the learner has no control over which parts of his information is private, and which is public.

On the other hand, there were concerns raised with regards to security. There exists literature, such as (Franz et al. 2006), on how to achieve basic security requirements: *confidentiality, integrity* and *access control*. The security of existing E-learning systems (such as Blackboard, WebCT, or Atutor) does provide a certain level of privacy. As such, **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. Moreover, **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 such that only the appropriate receiver can read the data. In the second case, **access control** mechanisms 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. Moreover, none of the previously mentioned security mechanisms can be used to observe the *core* of the definition of privacy, in such that the learner has no control on what information about him is being gathered by the E-learning system and how it is used. Although Privacy Policies have been provided for this purpose (Yee and Korba 2003), they cannot restrict unwanted access to the data.

Consequently, other approaches were proposed. (Anwar and Greer 2008) proposes a privacy mechanism based on identities. In particular, a learner can have different identities, or personas, that he could use within the different parts of the E-learning system. As long as the learner does not divulge his real identity, and the personas he is using are not linked to each other, or to the learner in question, his anonymity is insured, thus protecting his privacy. Another approach proposed in (Aïmeur et al. 2007) starts by proposing 4 different levels of privacy: *No Privacy*, *Soft Privacy*, *Hard Privacy* and *Full Privacy*. Each level of privacy protects different aspects of the learner's profile. Another dimension that is also considered, which is independent of the learner's personal data, is the tracking of learners within a course. Indeed, learners' activities within the system could be tracked, and a dossier could be built, even though their information within the system are protected. Hence, in addition to the privacy levels, (Aïmeur et al. 2007) also introduces 4 tracking level: *Strong Tracking*, *Average Tracking*, *Weak Tracking* and *No Tracking*. Each level of tracking reduces the amount of trace left by the learner within the E-learning system. In order to satisfy these various privacy and tracking levels, (Aïmeur et al. 2008) proposes *Anonymous Credentials for E-learning Systems* (ACES), a complete set of protocols, relying mainly on blind digital signatures and anonymous credentials, to preserve the learners' privacy.

23.3.3 Challenges

The previous sections presented why the need for privacy in E-learning, and highlighted several approaches to achieve that goal. Yet these existing approaches do

have their weaknesses, and this section details some of the major common shortcomings of the existing solutions, shortcomings that need to be addressed in order to have an effective privacy preserving E-learning.

One such weakness is the *overhead* produced by the privacy preserving mechanisms. Indeed, regardless of the chosen solution, preserving the privacy of the learner creates a computational and an operational overhead. Indeed, the cryptographic protocols used to protect the learner's privacy require significant amounts computational resources from the server hosting the E-learning platform, amounts that would grow with the increasing number of learners using the system. On the other hand, in order to preserve their own privacy, learners are required to perform additional operations. Indeed, privacy does come with a price, and learners are required to participate in the management of their information, whether by maintaining their identities, or their own anonymous credentials. Note that the higher the level of privacy required by the learner, the more complicated the privacy preserving approach will become, which implies a higher incurred overhead.

Another shortcoming of privacy is its impact on personalization. Indeed, most of the information gathered on learners within the E-learning system is used in order to personalize the learning experience, capitalizing on the learner's strength, while targeting his weaknesses, and thus tailoring the learning experience according to the learner's learning needs and style. Consequently, the personalization of the learning will be impacted by the fewer available information about the learner (due to his privacy preferences). Indeed, privacy and personalization are like two opposite forces *pulling* the learner's information: the first is pulling to hide it, whereas the second is pulling to gather more of it, in order to better personalize the content. It is a big challenge to find the middle ground such as to satisfy both the privacy and personalization needs.

Another aspect of privacy that requires further investigation is its *impact* on the learner. Indeed, although thus far, the existing research tends to demonstrate that privacy has a positive impact, to the best of our knowledge there were no studies conducted to evaluate the long impact of privacy on the learners. This lack of certitude, whether privacy has a positive or negative impact, is another weakness of privacy in E-learning. Indeed, privacy *might* provide this false sense of security: knowing that as long as you do enough to get the average and pass, no one will know. Consequently, learners might lose their motivation to perform, and they could become more nonchalant, or indifferent to the learning.

The challenges raised in this section relate to privacy preserving solution for E-learning systems. Nonetheless, the advent of what is commonly referred to as E-learning 2.0 raises a new set of challenges with respect to protecting learner privacy.

23.4 Privacy and E-Learning 2.0

E-learning 2.0 does not refer to a new class of LMS (Learning Management Systems) or a new educational technology. Rather it is a natural consequence of changes in how tutors and learners perceive learning in general. Indeed, in recent

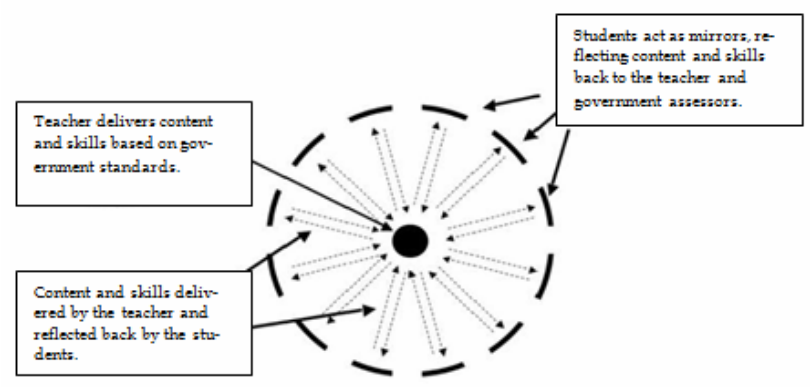

Fig. 23.3 Tutor-centered education (Webilus 2008)

years, education has been shifting from being *tutor-centered*, to being *learner-centered*. In tutor-centered education (Fig. 23.3), the tutor is the active participant in the educational process and learners are considered as passive receptacles of knowledge. Tutor-centered education is a *one size fits all* approach.

On the other hand, in learner-centered education (Fig. 23.4), the learners have access to a variety of knowledge sources and the tutor places more emphasis on what learners can contribute to the educational encounter.

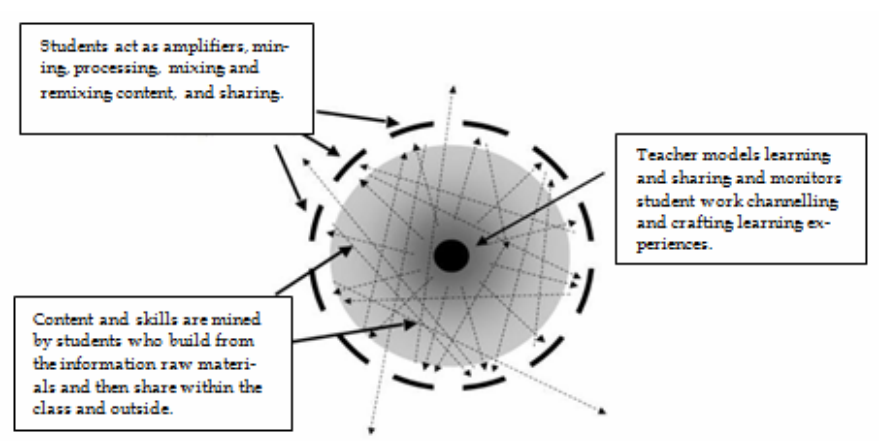

Fig. 23.4 Learner-centered education (Webilus 2008)

It is important to note that *E-learning 2.0 is not a consequence of Web 2.0*. Indeed, both share the same basic concept where the user/learner is not only a spectator and a simple consumer of information, but rather an active participant in the creation of such information. As such, one can view *Web 2.0* tools and technologies as a *natural recourse* to achieve *learner-centered* education.

We will start by defining Web 2.0, then highlight how Web 2.0 is used in E-learning, and then describe the impact on privacy and the challenges.

23.4.1 Web 2.0

Although the term Web 2.0 suggests a new version of the World Wide Web, it does not refer to an update or any technical specifications, but rather to changes in the ways software developers and end-users perceive and use the web. Indeed, the term Web 2.0 refers to a perceived second generation of web-based communities and hosted services (such as blogs, Wikis, etc.) which aim to facilitate creativity, and to promote collaboration and sharing between users.

In short, the following point summarizes the difference between Web 1.0 and Web 2.0: publishing vs. participation. Specifically, in Web 1.0 (*publishing*) the content is controlled by the publisher, and the users are just the recipient of the information. Whereas in Web 2.0 (*participation*) the users are no longer passive recipients of information, but are active participants in the creation of such information, participating in Wikis, tagging, rating, sharing, and/or referring websites. A recently published report (Lenhart et al. 2007) indicates that 64% of online teenagers in the US, ages 12 to 17, engage in at least one type of content creation.

There are three pillars to Web 2.0: the Social Web, Service Oriented Architecture (SOA) and Rich Internet Application (RIA).

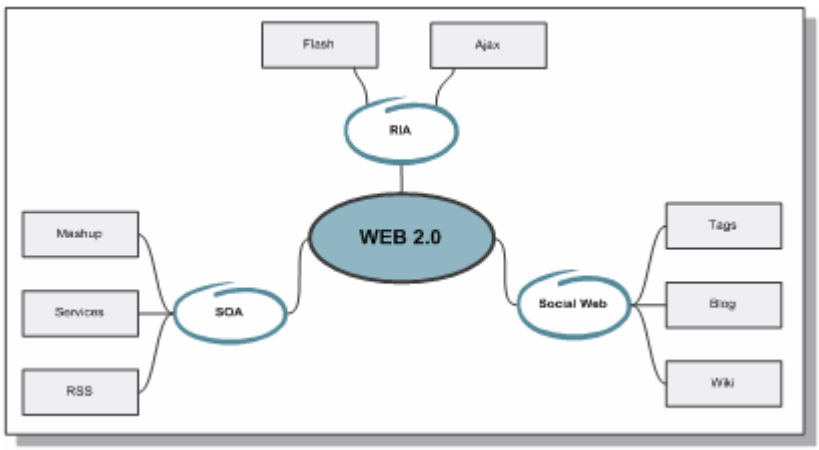

Fig. 23.5 The three pillars of Web 2.0, adapted from (Webilus 2008)

The **Social Web** refers to the "social interactions" between the users of the web, and the resulting virtual "social groups". It allows users to share their writings, videos, photos, and more with their friends, family, colleagues, or the public at large. For instance, the Social Web includes simple publishing through a blog or a wiki. As such, in the case of the blog the owner of the blog and his *faithful* readers can become a social circle where the readers can comment on the blog posts,

or each other's comments. Similarly, with the Wiki, the users who regularly visit, contribute to, or maintain the Wiki become a virtual social community centered on the Wiki.

The main drive behind the Social Web is collaboration and the harnessing of collective knowledge. Common features that exist in the Social Web, such as *tagging*, *rating*, *comments* and *recommendations*, exploit and share the knowledge and experiences of the users. As an example, we will consider social bookmarking sites, such as delicious.com or StumbleUpon.com. Such sites enable users to bookmark their favourite web sites, recommend and share these bookmarks with other users, or a community of friends.

Rich Internet applications (RIAs) are web applications that provide functionalities and interactions similar to desktop applications. Typically, RIAs are delivered through browser add-ons or directly through the webpage using for instance Ajax or Macromedia Flash. To illustrate RIAs, consider for instance Google documents (http://docs.google.com) which provides a decently complete set of tools to create and manage documents, spreadsheets, presentations and even forms. The whole set of tools is web based, that is accessible through the browser.

On the other hand, there is a multitude of web pages that illustrate the use of RIA, including web-based virtual computers, such as G.ho.st (http://g.ho.st/). Such environments provide a virtual computer environment, accessible online using any browser, which provides the functionalities and tools or a regular computer, including disk space (5 Gbytes in the case of G.oh.st), a media player, and even an office suite to create, and store documents spreadsheets and presentations.

Service Oriented Architecture (SOA) is an architectural style where the main goal is to relax the dependencies between various components and to achieve loose coupling. Specifically, a *service* is a task performed by the *service provider* to achieve a desired end result for a *service consumer*. Consequently, a service-oriented architecture is a collection of services (service providers and consumers), where these services communicate with each other. Such communication could be just simple data passing or it could involve two or more services coordinating to perform a certain activity. Note that the service provider can also be a service consumer. The flexibility and interoperability of SOA and web services has lead to a new type of web applications called Mashup. Specifically, a mashup describes a Web application that combines multiple services and/or data sources into one single application.

*23.4.2 Web2.0 and E-Learn*ing

This section highlights some examples of "Web 2.0" tools and websites designed for, and used in learning. For instance, a webcast consists of distributing media content over the using streaming media technology. A webcast may be distributed live or on demand. In essence, webcasting is "broadcasting" over the internet. A simple example of webcasting is a TV station that simultaneously streams over the internet the show being broadcasted on TV. On the other hand, a podcast is a series of media content made available via syndication, such as RSS. Dedicated software applications, known as podcatchers automatically identify and retrieve

new available media files. The utility of webcasts and podcast in E-learning is very clear: tutors can either webcast their lectures live to students, or the lectures could be made available on demand or through a podcast. Note that a lecture can consist of various media, such as audio only, a slide presentation with audio, a recording of the tutor, etc.

Currently, webcasting and podcasting are being used in several universities worldwide (Shim et al. 2007). It is important to note that webcasting and podcasting are not just used by virtual universities, but also as a complement to lectures in *traditional* classrooms, for instance, Berkely makes publicly available webcasts of several courses (available at http://webcast.berkeley.edu), consisting of either an audio recording of the tutor's lecture, a video recording of the tutor giving his lecture, or a slide presentation of the lecture with the explanations of the tutor.

Alternatively, wikis are websites that generally allow visitors at large to modify their content. Nonetheless, wikis generally can support authentication, such that certain members can modify only certain pages. This feature is important since it enable the use of wikis in group work assignments. Wikis offer the possibility of central access for all the users or limited user groups, which makes it an ideal choice for running projects, drafting documentations and other group work. As such, wikis are used to promote team work and collaboration between students (Raitman et al. 2005). Alternatively, wikis can also be employed by tutors to collaborate on creating learning content. For instance, wikiversity.org offers tutors the chance to collaborate and create freely available learning resources, where currently, on the English site of wikiversity, there are more than 10,000 pages available, covering various topics.

Similarly, *SuTree.com* and *eduSLIDE.net* offer both learners and tutors access to a variety of learning resources. Specifically, SuTree.com offers a variety of how-to videos, ranging from learning how to whistle, to following a complete course watching MIT lectures. eduSLIDE allows tutors to create lessons (presentations) and group them into courses, making these courses available for learners.

Additionally, many existing "web 2.0" pages and tools can help learners during the learning process. For instance, *Footnote.com* allows students to access primary source documents and photos, and to easily create and post online history reports. Moreover, *VoiceThread.com* can be used by both tutors (to create lessons) and learners (for homework purposes) to upload pictures and create an audio narrative to go along with them. *VisualThesaurus.com* offers, as its name indicates, a visual thesaurus. Specifically the lookup word is presented in the center of the graph, and edges connect the lookup word with its synonyms. A color code is used on the edge connecting the word to its synonyms to indicate whether the synonym is a noun, verb, adjective or an adverb. Moreover, the edge connecting the lookup word with its antonym is presented differently. *Wayfaring.com* is a mashup that uses Google maps to list podcasts and webcasts from about 68 universities worldwide. *wePapers.com* allows users to share academic papers, ranging from research papers, tutorials, lectures, to tests and exams. Moreover, users can comment, and even ask questions to the community about these papers. Another useful browser add-on is Diigo (http://www.diigo.com/). Diigo provides learners with the ability to highlight specific parts of webpages, add sticky notes and comments (private or

public) to the highlighted sections or the whole page, and learners can share the highlights and notes with their Diigo social network.

Moreover, existing systems rely on the learners' collaboration and social network to enhance the learning experience. For instance, *Knowledge Sea II* (Brusilovsky et al. 2005) treats research papers as regular pedagogical resources, allowing users to annotate and review these resources, and using the annotations to perform the recommendations. Comtella (Vassileva 2004) is another academic system that uses P2P (peer to peer) technology to enable students to share research papers. In addition, Comtella employs a *reputation* scheme (Mao et al. 2007) to motivate and award the students. On the other hand, SHAREK (Hage and Aïmeur 2008) is designed to enable learners to attach external learning resources to the tutor defined learning content within the E-learning system.

The proliferation of tools and websites such as listed earlier has led to the concept of Personal Learning Environment (PLE) (van Harmelen 2006). PLE is a combination of tools and processes, whether formal or informal, which learners use to gather information, reflect on it and work with it. The appeal of PLE for learners relies in the fact that they can choose the tools that best suit their preferences. An interesting representation we came once across compares a Learning Management System (LMS) and a Personal Learning Environment (PLE) using the following analogy: an LMS is similar to a Swiss army knife containing a set of tools, some of which you might never used. On the other hand, a PLE is like having a box containing the tools you use, but most importantly tools that you chose and prefer. Indeed, although it might be more practical to fit a large set of tools into your pocket (Swiss army knife analogy), having only the specialized tools that you are comfortable with does have it advantages.

Many PLE advocates portray an LMS as being inflexible and used to control the learning and the learner, whereas a PLE is portrayed as easy to use, personalized, and liberated. In short, LMS is equivalent to controlling how you learn, whereas PLE corresponds to giving you control over how you learn. Although controlled and passive learning reduces self reliance and causes loss of curiosity and creativity, an uncontrolled education would create a shortage of certified labor and would introduce unqualified people into the labor pool. Currently, this is where E-learning stands today (Fig. 23.6).

The Tutor delivers the learning content to the learner through the LMS. On the other hand, the learner has access to the controlled environment provided by the LMS as well as a PLE containing the set of his favorite tools and resources, which are external to the LMS. As such, the leaner can *freely* perform the learning activity, relying on the content and tools provided through the LMS, and on external *uncontrolled* resources through the PLE. In addition, the learner has access to both his personal social network (outside the LMS), and a peer network through the LMS. Note that some peers can also be part of the learner's external social network. In such a scenario, the tutor *controls* the curriculum (which courses and topics the learner must complete), and he can *validate* the learner's knowledge through assessments. On the other hand, the learner has the freedom to choose *how* to complete the learning activities: whether by solely using the content and

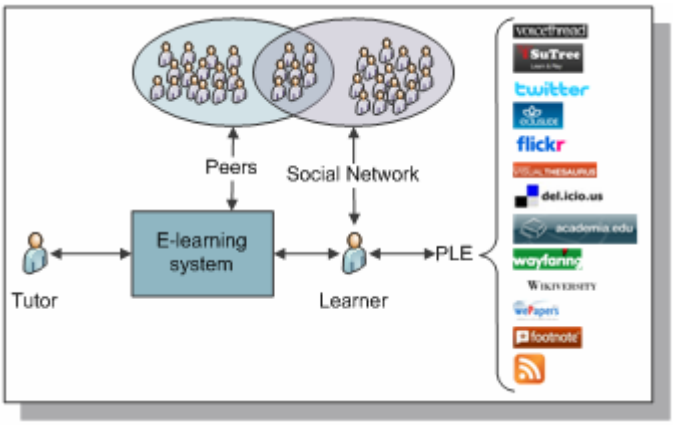

Fig. 23.6 Using LMS and PLE for education

tools provided through the LMS, by relying completely on his PLE, or a combination of both. In the last case, the LMS can be actually viewed simply as another component of the PLE.

Most of the PLE components are external to the E-learning system. Consequently, the educational institutions using an E-learning system, as well as the entity developing it, have no control over these components. This raises an enormous challenge with regards to ensuring and maintaining the learners' privacy.

23.4.3 Impact on Privacy and the Challenges

The challenges brought forth by the decentralized nature of PLEs with regards to learners' privacy are numerous. The first major concern is with regards to the security of these components. Indeed, as detailed in the section "*Pillars of security*", the Confidentiality and Integrity of the learner's information are imperative to protect his privacy.

Specifically, consider a learner using *delicious.com*, a social bookmarking website, to organize a list of websites and resources he is using to help prepare a report. If the confidentiality of that information is compromised, other learners – probably his classmates – might use the same resources in their report, effectively reducing his chances at a better grade. Similarly, the integrity of that information is also important: corrupting or altering these bookmarks would cause delays in preparing the report. Moreover, consider a learner using *zotero.org* to capture and organize references for a research paper he is working on. Again, the confidentiality, as well as the integrity of the information, is important.

The same need for confidentiality and integrity (not to mention availability) is necessary for any these components that are available for learning, and assuming that most do provide an acceptable level of security, would that be enough to protect the learners' privacy!? The answer is a resounding NO! Indeed, as with E-learning systems, the security is important factor to preserve privacy, but is not

enough. Specifically, in this case it is harder to protect the learners' privacy, because first, some of his information will be replicated across the various components he is using as part of his PLE. Indeed, most of these systems require registration, and ask for personal information, such as name, sex, email, etc. Having that information replicated all over the internet increases the risk of having parts of it, if not all, disclosed to unwarranted parties. Moreover, when using such systems, learners are providing large amount of information about themselves, and yet, are *blindly* trusting these systems not to disclose their information (Conti 2008). The use of solutions such as OpenID (openid.net) does help reduce the spread the learner's information across various systems, but it does not guarantee the *privacy* of the learner. Indeed, not all the components used for E-learning do support systems such as OpenID, and even when such support is available, the learner is not in control of his information, and he can still be easily tracked through the various systems.

23.5 Conclusion

Today, E-learning offers rich multimedia content and is more interactive. Moreover, E-learning is very flexible: students can choose instructor-led or self-study courses and they can select from a variety of learning tools that best fit their style. Indeed, one of the main advantages of E-learning is its adaptability to the learner's specific needs and preferences. Nonetheless, to do so, the E-learning systems collect large amounts of information about the learner information that could be misused, and therefore violating his *privacy.* The security of E-learning systems offers is imperative to safeguard the information stored within the system, and is essential to preserve privacy. Nonetheless, security alone is not enough and various solutions for privacy preserving E-learning were proposed, some relying on identities, others on anonymous credentials. Although these solutions are technically sound, they do fall short: they introduce a computational and operational overhead, influence the personalization of E-learning systems, and its effect (whether positive or negative) on learners attitudes is still not entirely determined. Preliminary research attributes a positive effect to privacy on learners, but this is a point that requires further investigation. Consequently, privacy preserving E-learning must be able to balance privacy, with access to learners' necessary information required to personalize the learning content and experience, while reducing the overhead incurred by privacy preserving mechanism. Alternatively, the advent E-learning 2.0 and the widespread use of PLEs introduced a new set of challenges that need to be addressed to ensure learner privacy. Indeed, learners regularly use and access resources external to the E-learning system or classroom. These resources are not controlled by the educational institution, and consequently are harder to supervise, increasing the risk to learner's privacy. Moreover, since most of these external resources require learners to register, their personal information will be redundantly duplicated, increasing the risk of unwanted disclosure of that information. Although not everybody will embrace our wish for privacy, as many

would agree, we consider privacy to be a fundamental human right: it is not negotiable! Since learning is as important, further research to concord privacy and E-learning is imperative.

References

Aïmeur, E., Hage, H., Mani Onana, F.S.: A Framework for Privacy-Preserving E-learning. In: Joint Itrust and PST conferences on Privacy, Trust Management and Security (IFIPTM 2007), Moncton (2007)

Aïmeur, E., Hage, H., Mani Onana, F.S.: Anonymous Credentials for Privacy-Preserving E-learning. In: The Montreal Conference on eTechnologies 2008, MCETECH 2008, Montreal (2008)

Anwar, M., Greer, J.: Role and Relationship-Based Identity Management for Private yet Accountable E-Learning. In: Joint iTrust and PST Conferences on Privacy, Trust Management and Security, IFIPTM 2008, Trondheim (2008)

Anwar, M., Greer, J.: Implementing Role-and Relationship-based Identity Management in E-learning Environments. In: 14th International Conference on Artificial Intelligence in Education, AIED 2009, Brighton, pp. 608–610 (2009)

BBC, Hackers hit Twitter and Facebook. BBC News (2009a), http://news.bbc.co.uk/2/hi/8188201.stm (Retrieved)

BBC, US man 'stole 130m card numbers' BBC News (2009b), http://news.bbc.co.uk/2/hi/americas/8206305.stm (Retrieved)

Blanchard, E., Roy, M., Lajoie, S., Frasson, C.: An evaluation of sociocultural data for predicting attitudinal tendencies. In: 14th International Conference on Artificial Intelligence in Education, AIED 2009, Brighton, pp. 399–406 (2009)

Brooks, C.A., Greer, J.E., Melis, E., Ullrich, C.: Combining ITS and eLearning Technologies: Opportunities and Challenges. In: Ikeda, M., Ashley, K.D., Chan, T.-W. (eds.) ITS 2006. LNCS, vol. 4053, pp. 278–287. Springer, Heidelberg (2006)

Brusilovsky, P., Farzan, R., Jae-wook, A.: Comprehensive personalized information access in an educational digital library. In: Proceedings of the 5th ACM/IEEE-CS Joint Conference on Digital Libraries, JCDL 2005 (2005)

Conati, C.: Probabilistic assessment of user's emotions in educational games. Journal of Applied Artificial Intelligence 16(7-8), 555–575 (2002)

Conti, G.: Googling Security: How Much Does Google Know About You? Addison-Wesley Professional, Reading (2008)

Dolog, P., Henze, N., Nejdl, W., Sintek, M.: Personalization in Distributed eLearning Environments. In: 13th World Wide Web Conference, New York, pp. 170–179 (2004)

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

Hage, H., Aïmeur, E.: Harnessing learner's collective intelligence: a Web2.0 approach to E-learning. In: Woolf, B.P., Aïmeur, E., Nkambou, R., Lajoie, S. (eds.) ITS 2008. LNCS, vol. 5091, pp. 438–447. Springer, Heidelberg (2008)

Hage, H., Aïmeur, E.: The impact of privacy on learners in the context of a web-based test. In: 14th International Conference on Artificial Intelligence in Education, AIED 2009, Brighton, pp. 65–72 (2009a)

Hage, H., Aïmeur, E.: The impact of privacy on learners in the context of a web-based test. In: 14th International Conference on Artificial Intelligence in Education, AIED 2009, Brighton (2009b)

IMS, IMS Global Learning Consortium, http://www.imsproject.org/ (Retrieved, September 2007)

Jerman-Blazic, B., Klobucar, T.: Privacy provision in e-learning standardized systems: status and improvements. Computer Standards & Interfaces 27(6), 561–578 (2005)

Lee, J.: Cyberattack rocks South Korea. GlobalPost (2009), http://www.globalpost.com/dispatch/south-korea/090710/cyberattacks (Retrieved)

Lenhart, A., Madden, M., Macgill, A.R., Smith, A.: Teens and Social Media (2007), http://www.pewinternet.org/PPF/r/230/report_display.aspU (Retrieved, January 2008)

LTSC Learning Technologies Standardization Committee (LTSC), http://www.ieeeltsc.org/ (Retrieved, September 2007)

Mao, Y., Vassileva, J., Grassmann, W.: A System Dynamics Approach to Study Virtual Communities. In: 40th Annual Hawaii International Conference on System Sciences, HICSS 2007, Waikoloa, Hawaii, p. 178a (2007)

Mccombs, B.L., Vakili, D.: A Learner-Centered Framework for E-Learning. Teachers College Record 107(8), 1582–1600 (2005)

Messmer, E.: DDoS attacks, network hacks rampant in oil and gas industry, other infrastructure sectors Network World (2010), http://www.networkworld.com/news/2010/012710-ddos-oil-gas.html (Retrieved)

O'Dell, J.: RockYou Hacker: 30% of Sites Store Plain Text Passwords The New York Times (2009), http://www.nytimes.com/external/readwriteweb/2009/12/16/16readwriteweb-rockyou-hacker-30-of-sites-store-plain-text-13200.html (Retrieved)

Raitman, R., Augar, N., Zhou, W.: Employing Wikis for Online Collaboration in the E-Learning Environment: Case Study. In: 3rd International Conference on Information Technology and Applications (ICITA 2005), Sydney, pp. 142–146 (2005)

Shim, J.P., Shropshire, J., Park, S., Harris, H., Campbell, N.: Podcasting for e-learning, communication, and delivery. Industrial Management & Data Systems 107(4), 587–600 (2007)

van Harmelen, M.: Personal Learning Environments. In: 6th International Conference on Advanced Learning Technologies (ICALT 2006), Kerkrade, pp. 815–816 (2006)

Vassileva, J.: Harnessing P2P Power in the Classroom. In: Lester, J.C., Vicari, R.M., Paraguaçu, F. (eds.) ITS 2004. LNCS, vol. 3220, pp. 305–314. Springer, Heidelberg (2004)

Webilus (2008), http://webilus.com/toutes-les-images (Retrieved, January 2008)

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

Woolf, B.P.: Building Intelligent Interactive Tutors, Student-Centered Strategies for Revolutionizing E-Learning. Elsevier & Morgan Kaufmann (2008)

Yee, G., Korba, L.: The Negotiation of Privacy Policies in Distance Education. In: IRMA International Conference, Philadelphia (2003)

Zeller, W., Felten, E.W.: Cross-Site Request Forgeries: Exploitation and Prevention (2008)

Zins, J.E., Bloodworth, M.R., Weissberg, R.P., Walberg, H.J.: The Scientific Base Linking Social and Emotional Learning to School Success. Journal of Educational and Psychological Consultation 17(2), 191–210 (2007)