Privacy Issues in Online Learning

Jim Greer

Professor of Computer Science
Director, University Learning Centre
Interdisciplinarity is like having one foot firmly planted in each of two canoes!
Motivation

• Learners have little privacy and are susceptible to identity crimes
• While there is a clear requirement for privacy, there is the necessity of data collection to award credentials and to provide personalized online learning
• Privacy and Trust hold a symbiotic relationship – privacy builds trust, and trust eases privacy concerns
• Online learning is an ideal milieu in which to investigate issues of privacy
Part I

- Privacy: What, Why, and How
What is Privacy?

• Privacy is a subjective, fluid ever-changing concept (Lessig, 1998)
  – the right to be let alone (Warren & Brandeis, 1890)
  – solitude (Brierley-Newell, 1998)
  – control over what someone does (Gavison, 1984)
  – freely behave without risk of being observed (Altman & Chemers, 1980)
  – context-dependent selective disclosure of personal information (Altman, 1975)
  – when, how, and to what extent information is communicated to others (Westin, 1967)

• Ability of individuals to control collection, retention, and distribution of personal information (Goldberg et al., 1997)
What is Privacy?

• Privacy is an interaction, in which the information rights of different parties collide (Noam, 1997)

• Privacy as a function of the monitored and the searchable (Lessig, 1998)

• Ability to reveal information selectively to negotiate social relationships most advantageous to the users (Rao & Rohatgi, 2000)

• About access control, data integrity, identity management
Why Privacy?

• Like freedom, we do not recognize its importance until privacy is taken away (David Faherty)

• protects us from being misdefined and judged out of context (Cavoukian & Hamilton, 2002)

• performs four functions for us: (Westin, 1967)
  – Threat to autonomy puts an individual under the control of those who know his secrets
  – Privacy provides moments “off stage,” when the individual can be: tender, angry, irritable, lustful, or dream-filled
  – privacy is essential for carrying on self-evaluation
  – Opportunity to share confidences and intimacies with a trustee—spouse, “the family,” friends, etc. (protected communication)
Privacy Protection

• a process of finding an appropriate balance between privacy and multiple competing interests (Song et al., 2006) — personalization, security, trust, etc.

• can be realized through access control and authentication

• implementation of sound security practice does not guarantee that privacy will be achieved (Menard, 2006)
  – The use of authentication when it is not needed could threaten privacy

• Privacy is generally approached as a social consideration, whereas security is seen as a technical concern
Trust

• choice to expose oneself to a risk toward one’s counterpart, in the expectation that the counterpart will not disappoint such expectation (Luhmann, 2000)
• a complex predictor of an entity’s future behavior based on the past evidence
• Assess trustworthiness to decide what piece of information would be safe, with whom, and in what context
• Use of trust is often implicit – a user who downloads a file from an unfamiliar web site trusts the web site implicitly
Privacy & Trust

- In the world, trust is used to manage privacy:
  - we share personal information with those we trust
  - we collaborate with someone trustworthy
  - we confide in someone trustworthy

- In the digital realm,
  - the level of self-disclosure depends on the user’s trust
  - Trust is influenced by the perceived level of privacy offered
Privacy & Trust

• Two important factors to build trust:
  – familiarity (Sheehan & Hoy, 2000), and
  – experiences (Doney & Cannon, 1997)

• To make the past experience positive, live up to the privacy and security expectations of the individual
Reputation-based Trust

- Reputation is a social notion of trust (Golbeck & Hendler, 2004)
  - we each maintain a set of reputations for the people we know
  - For an unknown person, we ask people with whom we already have relationships for information

- Though trust can be based upon many different sources (e.g. social rules, professional ethics, legal rules, etc.), reputation is the most effective source for measuring trust online

- In reputation-based trust, trust relationships grow based on longitudinal social behaviors of the actors
Trust, Privacy, Personalization

• Trust in the online transactions is closely related to issues of privacy (Friedman et al., 2000)
• In the online world, trust invokes the threat of privacy violation, identity theft, and threat to personal reputation
• In an asymmetric trust relationship, the weaker party trades privacy loss for a trust gain (Lilien & Bhargava, 2006)
  – How much of privacy is lost?
  – How much does a user benefit from certain trust gain?
• Trust promotes personalization: frank interactions $\rightarrow$ better data $\rightarrow$ better personalization
• Privacy and security can lead to trust (Andersson, 2005)
  – access control allows data disclosure only if the other party provides sufficient evidence of trustworthiness
  – secure access enhances users trust in the protection of personal data
• Privacy in the form of anonymity can diminish trust:
  – law enforcement difficult
  – individuals can behave in socially undesirable ways
  – cannot be sure who information is coming from (diminish integrity)
• Trust is both a pre-requisite and a consequence of good personalization practice (Briggs et al., 2004).
Privacy – Identity Management (IM)

- privacy is directly related to the knowledge of identity (Demchak & Fenstermacher, 2004)
- individuals reveal and conceal information selectively to maintain context-specific identity (Goffman, 1959)
- release all the information, but ensure the identities of the subjects are protected (Samarati, 2001; Sweeney, 2002)
- A teacher’s or learner’s privacy comes from their capacity to control the conditions were their identity information will be shared (Anwar, Greer, & Brooks, 2006)
- Contextual understanding and personal self-awareness help us properly control our identity and presentation during social interactions
Identity Management (IM)

- What is identity
  - A dataset of personal info (PI), (e.g. name, biometric information, behavioral pattern) used to model and thereby recognize an entity as distinct from others
  - An entity may be represented by many models (“partial identities”) including their own (“true identity”)

- Individuals hold multiple partial identity in multiple contexts:
  - A graduate student holds multiple partial identities based on the role they play: a student, a tutor, an instructor, or a marker

- Privacy is about protecting identity \( \rightarrow \) IM is a natural solution to privacy
IM : Anonymity or pseudonymity

• IM tools provide users with a desired control over their presentations of selves: anonymity, pseudonymity, or open identity.

• In anonymity, actions may occur and be observed but identity is not disclosed.

• Anonymity provides absolute privacy, but –
  – Does not allow personalization
  – Law enforcement difficult
  – Frees individuals to act in socially undesirable ways
  – Diminishes the integrity of information
  – Diminishes trust
Pseudonymity

- Selective PII is disclosed by associating a pseudonym with actions

- Allows us to attach longitudinal behavior records to an individual, but throws a blanket of secrecy over the true identity

- Supports reputation building (e.g. e-bay, slashdot, wikipedia)

- Good reputation gains trust, but reputation is associated with the pseudonym
Pseudonymity

• Digital identities and profiles are essential for personalization, but any kind of misuse causes violation of privacy, fraud, etc. (Mont et al., 2003)

• The pseudonym technology with unlinkability, accountability, and relative anonymity can give the user the ability to control the collection, retention, and distribution of personal information
Pseudonymity is effective, but—

- Explicit PII is hidden, but leaves information composite that may be fused to reveal full identify
- Need reputation transfer when merging/changing pseudonyms

- Information Giver (IG) should consider:
  - Benefit of revealing some PII
  - Intention of Information Seeker (IS)
  - What information Belongs to what context (as shown in context-specific identity model)
  - Sensitivity of each piece of information
Reputation Transfer (RT) model

• Anwar & Greer propose a model that supports
  – Generating contextual reputation
  – Vouching for a registered actor
  – Transferring reputation among multiple pseudonyms of an actor
• This model maintains privacy, but regulates bad actors
Reputation Transfer

• In RT model, there are 4 entities- actor, reputation, guarantor, and key-generator
  – An actor takes part in various e-learning activities
  – Reputation is the trustworthiness of an actor assessed over their past activities
  – A guarantor is a public actor who is a trusted witness of the past activities of a pseudonymous actor
  – A trusted key generator that facilitates Public Key Infrastructure
RT: Reputation Generation

• An actor registers its pseudonyms with a guarantor who would vouch for the actor and be credible in the community
• The guarantor periodically evaluates the reputation of the actor
• The actor gets an opportunity to contest any misrepresentation of their reputation to the guarantor
• The guarantor investigates the challenge and makes an appropriate adjustment to the reputation
RT: Vouching for an actor

• Upon request, a guarantor vouches for actors to their partners based on their reputation
• Occasional uses of anonymity can be facilitated by having trusted guarantor vouch for an actor using an anonymous identity
• A guarantor vouches for an actor in two ways:
  – i) responding to the queries about the actor
  – ii) responding to the actor’s reputation transfer request from one pseudonym to another
RT: Transferring Reputation

• The key-generator (KG) generates key-pair for both guarantor and actor
• The guarantor publishes the public key so that any actor can make encrypted service requests
• An actor registers with the guarantor – register (pseudonym, public-key)
• The guarantor gives a unique registration number (RN) for each of its clients
• The guarantor generates and finalizes reputation
• KG generates another key-pair for reputation
RT: Transferring Reputation

0. Generate Key Pair
1. Send Public Key: GPub
2. Encrypted-by-GPub(Register(pseudonym, APub))
3. Generate-Reputation(pseudonym)
5. Encrypted-by-APub(Send(Reg#, Reputuation))
6. Encrypted-by-GPub(Contest/Accept Reputuation)
7. Generate Reputation Key Pair
8. Send Reputation Key Pair
9. Encrypted_by_RPub(reputation)
10. Encrypted-by-APub(Send(RPri, RD, Reputation))
11. Decrypted_by_RPri(reputation)
12. Transfer to(reg#, RD)
13. Transfer from(reg#)

Follow step 7-10 & reputation transfer is done

RD: Reputation Digest
RPub: Public Key to encrypt reputation
RPri: Private Key to decrypt reputation
GPub: Guarantor’s public key
APub: Actor’s public key

ARIES Lab, Computer Science, University of Saskatchewan
RT: Transferring reputation

• The guarantor sends the reputation private key, R_Pri and reputation certificate to the actor
• The guarantor also generates a reputation digest, R_D (MD5 hash) to uniquely identify each certificate
• A reputation transfer is a two way process that has to be initiated by the transferring actor and followed by the receiving actor
• First, the transferring actor makes a request to the guarantor providing the receiving actor’s secret R_N and R_D
Transferring reputation

- Then the receiving actor makes a similar request by providing the transferring actor’s RN
- RD is checked to confirm a valid reputation transfer request
- Upon verifying non-repudiation, the guarantor requests a new reputation key pair from KG
- the guarantor encrypts transferring actor’s reputation with the new public key and sends the corresponding RPri, RD, and reputation to the receiver
Private Transfer

• Since both the transferring and receiving actors are registered users of a guarantor, any bad acting can be traced and verified.
• All the communication between an actor and the guarantor takes place using each other’s public key.
• The integrity of reputation can be checked using the reputation digest.
Part 2

• Privacy in Online Learning

• Privacy in Social Networking
Online learning

• The realm of online learning has been expanded from academia, to industry, to cyber communities
• Due to participation and personalization, the Read-Web has transformed into a Read-Write-Web
• Online learning has become a personal learning center, where content is reused and remixed to cater to the students own needs and interests (Downes, 2005)
• Besides institutional or corporate learning, informal learning widely takes place in personal networks, or in the communities of practice
Online learning

- In most part, online learners' peers (and instructors) are strangers
- Interactions are often devoid of visual and verbal cues
- As a result, personalization is not possible without explicit learner modeling
- Lack of contextual cues – any interaction can be taken out of context and used to misrepresent the interlocutor
- learners should be able to separate acting within the online learning environment from other roles in their life
Personalization in online learning

• Personalization of learning involves the presentation of a learning experience that is customized to the preferences of the learner (Dagger et al., 2003).

• Personalization is a need in online learning, because of:
  – diverse learning objects,
  – different cognitive abilities,
  – different level of prerequisites, and
  – different learning styles (Borcea et al., 2005)
Personalization

- personalization can be based on multiple paradigms (Dagger et al., 2003):
  - Context personalization is adapting to the preferences of the learner
  - Competency personalization is adapting to the learner’s prior knowledge
  - Prerequisite personalization is adapting to the currently required prerequisites of the learner

- learner modeling is the prerequisite for personalization
Personalization

• learner modeling involves: understanding the world view of the learner and tracking information about a learner (McCalla, 2000)

• "trust" is an important concern in elearning systems (Xu & Korba, 2002)
  – service provider must trust that a learner truly has legitimate credentials and authorization to participate
  – the learner must trust that the service provider will use private information appropriately
Personalization in Online Learning

• Learners typically have trust in the system

• An effective collaboration, whether synchronous (e.g. chat, conferencing) or asynchronous (email, blogs, threaded discussions), depends upon trust

• The primary privacy concern in collaborative work and collaborative learning is “impression management” (Patil & Kobsa, 2003)

• In online learning, little consideration is given to privacy and security
Privacy in e-learning

• Why privacy in online learning?
  – Learner rights
  – Establish an unbiased environment (Borcea et al., 2005)
  – Learner and instructor comfort
Privacy in e-learning

• Since an online learning application aims at assisting users, they cannot act in full anonymity (Borcea et al., 2005)

• Two aspects of personal data that pertain to privacy protection of learners (Borcea et al., 2005; Franz et. al, 2006):
  – Data parsimony: store as little personal data as possible
  – Data partitioning: partition data into context-specific partial identities
Social Networks

• The social networking sites (SNS) are another frontier for exploring learner’s privacy

• Facebook, MySpace, Friendster
  – Millions of users
  – Gives the perception of the online space as a closed, trusted, and trustworthy community

• “Finding oneself” vs “Inventing oneself”
Social Networks

• SNS provides a fertile ground for identity development and cultural integration

• Individuals give away their “profile” and their own social networks

• Revealing personal information to vast networks of loosely defined acquaintances and complete strangers should be worrisome!!!

 Dochek
Social Networks

• In many occasions we want to share information with a small circle of close friends, and not strangers

• In other instances we want to share personal information with strangers, but not with those who know us better

• People are indicated as Friends even though the user does not particularly know or trust the person
  – In Facebook interactions, “there is no way to determine the metric used or the role or weight of the relationship” (Boyd, 2004)
Social Networks

• The factors that drive information revelation (Acquisti, 2004):
  – Peer pressure and herding behavior
  – relaxed attitudes towards (or lack of interest in) personal privacy
  – incomplete information (about the possible privacy implications of information revelation)
  – faith in the networking service or trust in its members
  – myopic evaluation of privacy risks

• provide opportunities to combine online and face-to-face interactions
Social Networks

• college-oriented networks offer a wealth of personal data to external observers
  – the Pentagon manages a database of 16-to-25-year-old US youth data, containing around 30 million records (Cave, 2005)
• Privacy expectations may not be matched by privacy reality
• By default, everyone in Facebook appears in searches of everyone else
  – users tend to not change default settings – Mackay, 1991)
Growing an Identity

• Blogging provides new ways for people to engage in self-expression and self-development

• Digital expressions have properties not normally considered in everyday life; they are easily copied, searched, or archived

• In digital environments, the lack of presence makes it difficult to know who is listening. Thus, how are unknown audiences negotiated? (Danah & Heer)
  – Fakesters in Friendster!!
Privacy in Social Networks

• Often, the topic of persistent conversations raises critical privacy issues. What happens when your future boss accesses your information? What happens when a big company buys your data? What happens when your social network is modelled?

• Public online is very different form public offline: persistence and exact copies are not something that people think to negotiate when they think about the nature of being public.

• Teens and adults have developed different notions of privacy: young people feel relatively comfortable sharing aspects of their lives.
Part 3

Recommendations for building privacy-enhanced and personalized online learning environments
Traditional Classroom: Privacy & Personalization

- Traditional classroom represents a closely knit group
- Yet some information is protected
- Physical presence warrants authenticity
- Instructor provides some degree of personalization by observing the visual cues of learners
- Contextual cues are available – teachers might say provocative things for pedagogical purpose
Recommendations in building privacy-enhanced online learning environment

• **Allow pseudonymity**: Allow users to use self-picked/system generated identifiers when divulging part of their context appropriate identity

• **Allow anonymity** when possible: Allow users to perform low risk activities anonymously

• **Facilitate information sharing** based on trust: help users evaluate the trustworthiness of other actors and allow them to perform some degree of social commitment
Recommendations

• Allow attaching **contextual cues**: role-based access to information

• Allow attaching **verbal and non-verbal cues** with information: by providing tags or emoticons

• **Detect/purge** unnecessary personal information: warn users about **privacy slips**
Recommendations

- Allow information to expire: Attach “time to live” tags for each piece of information

- Promote privacy awareness: educate users about privacy and the risk of identity disclosure

- Punish bad actors: flag bad behaviors and recognize the good users with a higher reputation score
iHelp

• An e-learning system which doubles as a research platform in advanced ed. tech.
  – iHelp Courses (LMS)
  – iHelp Discussion (forum)
  – iHelp Chat (cohort-based chatrooms)
  – iHelp Share (messaging and co-annotation)
  – iHelp Lecture (video lecture capture)

• Detailed tracking and monitoring for personalization and research
Privacy in iHelp

- The iHelp system supports privacy
  - Role-based access control
  - Closed cohorts and groups
  - Pseudonymity: allows multiple pseudonyms – system default or user picked
  - Anonymity: instructors may enable the option of anonymous posting
  - Context Separation: provides context specific interaction channels
Privacy in i-Help

- **Facilitation of Trust**: As learners interact with one another, familiar pseudonyms emerge and attribution of personalities to pseudonyms quickly develops.

- **Detection and removal of unnecessary personal information**: provides aggregate information to instructors, learners and researchers stripping off personal information where appropriate.
Privacy in i-Help

- **Promoting privacy awareness**: the system presents its privacy policy to its users and raises privacy awareness

- **Punishment for bad actors**: administrators can trace the true identity of a pseudonymous or anonymous user, but only if need be
Conclusion

• **Some recommendations** are made in building privacy-enhanced and personalized online learning environments.

• **Implementation** of some of these recommendations in iHelp concludes that it provides a reasonable degree of privacy protection for learners, facilitates trust, and allows personalization.

• **A reputation system** (that transfers/merge reputation across multiple pseudonyms) together with an IM system facilitate both privacy & trust.

• **We plan to implement a protocol for information expiration** that can be implemented within our systems using the time-to-live tag for each piece of information.
Acknowledgements

• NSERC
• Mohd Anwar, Chris Brooks, Zinan Guo
• Other members of the ARIES Lab

• Contact me
• Jim.Greer@usask.ca