Technical Solutions for Privacy-Enhanced Personalization

Yang Wang
Bren School of Information and Computer Sciences
University of California, Irvine
Donald Bren Hall 5091
Irvine, CA 92697-3440
yangwang@uci.edu
(949)-351-9336 (phone)
(949) 824-4056 (fax)

Alfred Kobsa
Bren School of Information and Computer Sciences
University of California, Irvine
Donald Bren Hall 5092
Irvine, CA 92697-3440
kobsa@uci.edu
(949) -202-5704 (phone)
(484)-762-6644 (fax)

To appear in Constantinos Mourlas and Panagiotis Germanakos, eds.: Intelligent User Interfaces: Adaptation and Personalization Systems and Technologies. Hershey, PA: IGI Global.

Technical Solutions for Privacy-Enhanced Personalization

ABSTRACT

This chapter presents a first-of-its-kind survey that systematically analyzes existing privacy-enhanced personalization (PEP) solutions and their underlying privacy protection techniques. The evaluation is based on an analytical framework of privacy-enhancing technologies, an earlier work of the authors. More specifically, we critically examine whether each PEP solution satisfies the privacy principles and addresses the privacy concerns that have been uncovered in the context of personalization. The chapter aims at helping researchers better understand the technical underpinnings, practical efficacies and limitations of existing PEP solutions, and at inspiring and developing future PEP solutions by outlining several promising research directions based on our findings.

KEYWORDS

Online privacy, privacy laws, privacy regulations, hypermedia, security, personalization, privacy-enhanced personalization, privacy-enhancing technology

INTRODUCTION

Privacy and personalization are currently at odds (Kobsa, 2002, 2007ab; Teltzrow & Kobsa, 2004; Wang & Kobsa, 2006). For instance, online shoppers who value that an online bookstore can give them personalized recommendations based on what books they bought in the past may wonder whether their purchase records will be kept truly confidential in all future. Online searchers who are pleased that a search engine disambiguates their queries and delivers search results geared towards their genuine interests may feel uneasy that this entails recording all their past search terms. Students who appreciate that a personalized tutoring system can provide individualized instruction based on a detailed model of each student's understanding of the different learning concepts may wonder whether anyone else besides the system will have access to these models of what they know and don't know.

Various technical solutions have been proposed to safeguard users' privacy while still providing satisfactory personalization, e.g., on web retail or product recommendation sites. Technical solutions for privacy protection represent a special kind of so-called Privacy-Enhancing Technologies (PETs). In (Wang & Kobsa, forthcoming), we propose an evaluation framework for PETs that considers the following dimensions:

(1) What high-level principles the solution follows
We identify a set of fundamental privacy principles that underlie various privacy laws and regulations and treat them as high-level guidelines for enhancing privacy.

(2) What privacy concerns the solution addresses
We analyze privacy solutions along major privacy concerns that were identified in the literature.

(3) What basic privacy-enhancing techniques the solution employs
We look at the technical characteristics of privacy solutions, to critically analyze their effectiveness in safeguarding privacy and supporting personalization.

The rest of this chapter is organized as follows. Firstly, we describe and categorize major privacy principles from privacy laws as well as other desirable principles in the context of privacy protection

(we thereby largely follow (Wang & Kobsa, forthcoming)). Secondly, we discuss privacy concerns and how different privacy principles address them. Thirdly, as the central contribution of this chapter, we describe the techniques that have been used in the main types of privacy-enhancing personalization solutions, and how they relate to the major privacy concerns and privacy principles. Fourthly, we discuss findings from this analysis. Finally, we collude with future research directions.

PRIVACY PRINCIPLES

Privacy legislation and regulation is usually based on more fundamental privacy principles. In our framework, we select a comprehensive set of major principles from our survey of over 40 international privacy laws and regulations (Kobsa, 2007b; Wang, Zhaoqi, & Kobsa, 2006). Any principle manifested in these privacy laws and regulations was included in our framework if it has impacts on how web-based personalized systems operate. Besides, we also define or identify other principles/properties that are desirable for privacy enhancement and personalization. Additional principles may possibly need to be added in the future, as new personalization technologies with new privacy threats emerge or the concept of privacy evolves. Below we list our principles, grouped by their provenance.

Privacy principles from privacy laws, regulations and recommendations

1. Notice/Awareness
   - Clarity: *Make these privacy policy statements clear, concise, and conspicuous to those responsible for deciding whether and how to provide the data* (Kobsa, 2007b; USACM, 2006);
   - Notice upon collection: *Whenever any personal information is collected, explicitly state:*
     - *the precise purpose of the collection,*
     - *all the ways in which the information might be used,*
     - *all the potential recipients of the personal data,*
     - *how long the data will be stored and used;*
2. Minimization
   Before deployment of new activities and technologies that might impact personal privacy, carefully evaluate them for their necessity, effectiveness, and proportionality: the least privacy-invasive alternatives should always be sought (USACM, 2006).
3. Purpose specification
   The purposes for which personal data are collected should be specified not later than at the time of data collection and the subsequent use limited to the fulfillment of those purposes or such others as are not incompatible with those purposes and as are specified on each occasion of change of purpose (OECD, 1980).
4. Collection limitation
   There should be limits to the collection of personal data and any such data should be obtained by lawful and fair means […] (OECD, 1980).
5. Use limitation
   Personal data should not be disclosed, made available or otherwise used for purposes other than those specified (OECD, 1980).
6. Onward transfer
   Personal data should not be transferred to a third country/party if it does not ensure an adequate level of protection (EU, 1995; FTC, 2000c)
7. Choice/Consent
   Where appropriate, individuals should be provided with clear, prominent, easily understandable, accessible and affordable mechanisms to exercise choice in relation to the collection, use and disclosure of their personal information (APEC-FIP, 2004). The two widely adopted mechanisms are:
   - Opt-in: *requires affirmative steps by the consumer to allow the collection and/or use of information* (FTC, 2000a);
   - Opt-out: *requires affirmative steps to prevent the collection and/or use of such information* (FTC, 2000a).

8. Access/Participation
   An individual should have right to:
   - *know whether a data controller has data relating to her* (OECD, 1980),
   - *inspect and make corrections to her stored data* (USACM, 2006)
9. Integrity/accuracy
   A data controller should ensure the collected personal data is sufficiently accurate and up-to-date for the intended purposes and all corrections are propagated in a timely manner to all parties that have received or supplied the inaccurate data (USACM, 2006).
10. Security
    Personal data should be protected by reasonable security safeguards against such risks as loss or unauthorized access, destruction, use, modification or disclosure of data (OECD, 1980).
11. Enforcement/Redress
    Effective privacy protection must include mechanisms for enforcing the core privacy principles. At a minimum, the mechanisms must include (FTC, 2000b):
    - Recourse mechanisms for customers: *readily available and affordable independent recourse mechanisms by which an individual's complaints and disputes can be investigated and resolved and damages awarded where the applicable law or private sector initiatives so provide;*
    - Verification mechanisms for data controllers: follow-up procedures for verifying that the attestations and assertions businesses make about their privacy practices are true and that privacy practices have been implemented as presented;
    - Remedy mechanisms: obligations arising out of failure to comply with these principles by organizations announcing their adherence to them, and consequences for such organizations.

Anonymity-related principles from the security literature

12. Anonymity
    Anonymity means that users cannot be identified nor be tracked online.
13. Pseudonymity
    Pseudonymous users also cannot be identified, but can be tracked by their unique "aliases" or "personae".
14. Unobservability
    A data controller cannot recognize that a system/website is being used/visited by a given user.
15. Unlinkability
    A data controller cannot link two interaction steps of the same user.
16. Deniability
    Deniability means that users are able to deny some of their characteristics or actions (e.g., a visit to a particular website), and others cannot validate the veracity of this denial.

Other desirable principles for privacy enhancement, mostly from human-computer interaction research

17. User preference
    Different users can have different privacy preferences. A data controller should tailor its privacy practices to each individual user's preferences.
18. Negotiation
    This principle calls for the support of negotiation between users and websites so that they can agree on the privacy practices that the website can follow.
19. Non-intrusiveness
    Non-intrusiveness means that users have control over incoming information. Popup ads and junk emails are typical example for intrusiveness.
20. Ease of adoption
    This principle considers how easy it is for organizations to implement a given privacy protection solution, for instance, whether the solution relies on special or unusual protocols or proprietary technologies, or on technologies that are not readily available.

21. Ease of compliance
    An increasing number of legal privacy duties have been imposed on data controllers, such as to monitor and provide audit trails of their factual privacy practices. This principle is concerned with the ease of meeting such legal requirements by adopting a specific privacy protection solution.
22. Usability
    A privacy protection solution should be easy on users, e.g., user involvement should be reasonable.
23. Responsiveness
    The privacy protection solution should respond promptly to changes of a user's privacy decisions.

Desirable principles for personalization

24. Personalization quality
    This principle is concerned with maximizing the personalization quality and associated benefits.

PRIVACY CONCERNS

There exist various approaches to categorize privacy (Camp & Osorio, 2003; Solove, 2006; Wang, Lee, & Wang, 1998), and they seem to have three main themes in common: the protection of people's identities, people's right to seclusion, and their right to control their data (such as to decide what data can be collected or disclosed for what purpose, how their data will be used, with whom the data may be shared, etc.). In Table 1, we categorize the 24 identified principles by the type of privacy protection that they afford. Notice that the general category contains principles that afford all three types of privacy protection.

A web personalization process is typically comprised of three tasks (Kobsa, Koenemann and Pohl, 2001):

1 Acquisition
  This task involves: (1) gathering information about users' characteristics, computer usage behavior and the usage environment, and (2) building a user model, a usage model and an environment model.
2 Representation and secondary inference
  This task consists in expressing the content of the user model and usage model in a formal system, allowing further access and processing.
3 Production
  This task is concerned with the adaptation of content, presentation, modality and structure of information conveyed to the user, based on the user, usage and environment models.

Another way of understanding web personalization is to dissect it in terms of higher-level system activities that it may entail, such as tracking user interactions with websites, creating user profiles based on the interaction logs, generating personalized recommendations to users based on their logs and profiles, and contacting users with personalized recommendations for potential purchases. These activities may cause different privacy concerns at varying degrees of likelihood. For instance, sharing users' personal data with third parties will be *very likely* to cause concerns over improper transfer of personal data, while it will be *likely* to engender concerns over unwanted solicitation (e.g., that third parties use the shared personal information to advertise their products to them).

Wang et al. (H. Wang, Lee, & Wang, 1998) present a taxonomy of privacy concerns in Internet marketing including improper access, improper collection, improper monitoring, improper analysis, improper transfer, unwanted solicitation and improper storage. These concerns as well as improper merge (of data) also seem to apply to web personalization. Table 2 that is based on (Teltzrow & Kobsa, 2004; Wang, Lee, & Wang, 1998) shows what privacy concerns (columns) are *very likely* or *likely* to arise from web personalization activities (rows). Table 3 depicts what privacy concerns (columns) might be involved in the tasks of a web personalization process (rows).

Table 1. Categorization of principles based on the type of privacy protection

Privacy / **Principle**	General	Protection of Identity	Seclusion	Control over data
Notice/Awareness	X			
Minimization				X
Purpose specification				X
Collection limitation				X
Use limitation				X
Onward transfer				X
Choice/Consent	X			
Access/Participation				X
Integrity/accuracy				X
Security				X
Anonymity		X		
Pseudonymity		X		
Unobservability		X		
Unlinkability		X		
Deniability		X		
Enforcement/Redress	X			
User preference	X			
Negotiation	X			
Seclusion			X	
Ease of adoption	X			
Ease of compliance	X			
Usability	X			
Responsiveness	X			
Personalization quality	X			

Table 2. Potential privacy concerns in potential web personalization activities

	Control over data							Seclusion	Protection of identity
	Improper acquisition			Improper use			Improper storage	Unwanted solicitation	Identity fraud/theft
	Improper access	Improper collection	Improper monitoring	Improper analysis	Improper merge	Improper transfer			
Tracking		XX	XX						
Profiling		X	X	X	X	X			X
Cross-website recommendation		X	X	X	XX	XX	X	X	
Single-website recommendation		X	X	X	X	X	X	X	
Third-party data sharing				XX	X	XX	X	X	X
Direct mailing				X				XX	

XX: Very likely X: Likely

Table 3. Potential privacy concerns in web personalization proces

	Control over data							Seclusion	Protection of identity
	Improper acquisition			Improper use			Improper storage	Unwanted solicitation	Identity fraud/theft
	Improper access	Improper collection	Improper monitoring	Improper analysis	Improper merge	Improper transfer			
Acquisition	X	X	X			X	X		X
Representation & secondary inference				X	X		X		X
Production				X	X	X		X	

TECHNICAL SOLUTIONS FOR PRIVACY-ENHANCED PERSONALIZATION

Our framework for evaluating the effectiveness of technical solutions for safeguarding privacy whilst supporting meaningful personalization assesses privacy solutions along three different dimensions: (1) what high-level privacy principles the solution follows, (2) what privacy concerns it addresses, and (3) what basic privacy-enhancing techniques it employs. In the preceding sections, we identified 24 major principles and 3 major privacy concerns and presented their relationship to each other. In this section we discuss the major privacy-enhancing personalization solutions that have been proposed today, what basic privacy-enhancing techniques they employ, and how these solutions relate to the described principles and privacy concerns.

Pseudonymous personalization

Pseudonymous personalization allows users to remain anonymous with regard to the personalized system and the whole network infrastructure, whilst enabling the system to still recognize the same user in different sessions so that it can cater to her individually. Most of these techniques allow a user to have more than one pseudonym/account/role/persona, so that the user can keep apart different aspects of their online activities (e.g., work versus entertainment).

The Janus Personalized Web Anonymizer (Gabber, Gibbons, Matias, & Mayer, 1997) serves as a proxy between a user and a web site. For each distinct user-website pair, it utilizes a cryptographic function to automatically generate a different alias (typically a user name, a password and an email address) for establishing an anonymous account at the website. Janus also supports anonymous email exchanges from a website to a user, and filters the potentially identifying information of the HTTP protocol to preserve user privacy.

Arlein et al. (Arlein, Jai, Jakobsson, Monrose, & Reiter, 2000) suggest an infrastructure that enables global user profiles to be maintained and accessed by different merchants. Users can control their data disclosure by grouping their information into profiles pertaining to different personae and can selectively authorize merchants to access these profiles. The infrastructure includes a persona server to assist users manage their personae. The persona server is separate from the profile database, so as to prevent linking different profiles of the same user. Besides, the infrastructure also has a tainting-based access control mechanism that allows merchants to designate which data about user interaction at their sites can be accessed by other merchants.

Ishitani et al. (Ishitani, Almeida, & Wagner, 2003) implemented a system called Masks (Managing Anonymity while Sharing Knowledge to Servers). The system consists of both server-side and client-side components, namely the Masks server and the privacy and security agents (PSAs). The Masks server, acting as a proxy between users and websites, manages masks (temporary group identifications that are associated with specific topics of interest) and assigns them to users. This enables user information to be collected under those masks and enables the users to receive group-based personalization. The PSAs runs with users' web browsers and allows users to configure the masks as well as other functionalities such as blocking and filtering cookies and web bugs.

Kobsa and Schreck (Kobsa & Schreck, 2003) propose a reference architecture for pseudonymous yet fully personalized interaction. The architecture includes a MIX network between applications and user modeling servers, supports standard anonymization techniques between clients and applications, offers a choice of encryption at the application and the transport layers, and a hierarchical role-based access control model. One privacy enhancement of this architecture over other anonymization or pseudonymization techniques is that it hides both the identities of the users and the location of the user modeling servers in the network.

Hitchens et al. (Hitchens, Kay, Kummerfeld, & Brar, 2005) present an architecture that allows users to easily create their personas (a subset of a user model), and to selectively share these authenticated pseudonymous personas with certain service providers (via user defined preferences). Service providers can use the information contained in the personas to tailor their services to users.

Table 4 presents an analysis of the aforementioned pseudonymous personalization systems along the following characteristics:

Table 4. Pseudonymous personalization systems and their characteristics

System / Characteristics	Janus	Global user profile infrastructure[1]	Masks	Pseudonymous personalization reference architecture[2]	Personas architecture[3]
GENERAL					
Alias-to-website cardinality	1:1	m:n	m:n	m:n	m:n
User control		+	+	+	+
Personalization	Single site single user	From single site single user to cross-site single user	Group based	Cross-site single user	From single site single user to cross-site single user
PROCEDURAL ANONYMITY					
Sender/user anonymity	+	+	+	+	+
Receiver/website anonymity		+			
UMS anonymity				+	
CONTENT-BASED ANONYMITY					
Content-based anonymity				+	
LINKABILITY					
Linkability for a single pseudonym	+	+	+	+	+
Unlinkablity of pseudonyms for a user	+	+	+		+

+: Support

[1] (Arlein, Jai, Jakobsson, Monrose, & Reiter, 2000; Kobsa, 2002, 2007b; Teltzrow & Kobsa, 2004)

[2] (Kobsa, 2007b; Kobsa & Schreck, 2003)

[3] (Hitchens, Kay, Kummerfeld, & Brar, 2005; Kobsa, 2007b)

1. Alias-to-website cardinality
   The alias-to-website cardinality describes the relationship between the number of aliases pertaining to a user and the number of websites at which the alias(es) may be used. For example, a cardinality of 1:1 means that each user will have exactly one alias for every website, while 1:n means that a user has one global alias/profile for all websites, and m:n means that a user can have an arbitrary number of aliases for any number of websites.
2. User control
   User control denotes whether the system allows users to control the usage of their alias/profile at different websites.
3. Personalization
   This factor evaluates to what extent the websites can provide personalized services to users. For example, a site can provide personalized services using the user's interaction logs with this site, or it could use the logs from multiple sites.
4. Sender anonymity
   Sender anonymity indicates whether or not users are identified in the interactions.

5. Receiver anonymity
   Receiver anonymity indicates whether websites are identified in the interactions.
6. User Modeling Server (UMS) anonymity
   UMS anonymity indicates whether or not user modeling servers (or more general, the repositories that store the user models/profiles) are kept anonymous.
7. Content-based anonymity
   Content-based anonymity prevails when no identification by means of the exchanged data is possible.
8. Linkability for a single pseudonym
   This characteristic indicates whether or not a user's interaction steps or sessions with one or multiple websites can be linked using one pseudonym of hers
9. Unlinkability of pseudonyms for a user
   This characteristic indicates whether or not multiple pseudonyms pertaining to the same user can be linked.

At first sight, pseudonymous personalization seems to be a panacea for all privacy problems because it seems to protect identity and, in most cases, privacy laws do not apply any more when the interaction is anonymous. However, anonymity is currently difficult and/or tedious to preserve when payments, physical goods and non-electronic services are being exchanged. It harbors the risk of misuse, and it hinders vendors from cross-channel marketing (e.g. sending a product catalog to a web customer by mail). Besides, users may still have additional privacy preferences such as not wanting to be profiled even when done pseudonymously only, to which personalized systems need to adjust. Moreover, Rao et al. (Rao & Rohatgi, 2000) point out that pseudonymity, or more broadly, hiding explicit identity information (e.g., name, email address) is not sufficient to guarantee privacy. They demonstrate using a technique from stylometry (a field of linguistics that uses syntactic and semantic information to ascribe identity or authorship to literary works), and principal component analysis of function words, to attack pseudonymity. Similar findings were made for of database entries (Sweeney, 2002), web trails (Malin, Sweeney, & Newton, 2003), query terms (Nakashima, 2006), and ratings.

Distributed personalization

Distributed personalization for safeguarding users' privacy has so far primarily been investigated in the domain of collaborative filtering (CF). Collaborative filtering is a popular technique for generating personalized recommendations using other users' preferences. The underlying assumption is that a user will prefer things that similar users like. In general, CF techniques use weighted combinations of nearest neighbor ratings to make predictions based on a user's preferences. A number of algorithms exist to determine proximity, including correlation between users, vector similarity methods, Bayesian clustering and Bayesian networks.

In recommender systems based on CF techniques, distribution may affect two aspects: the storage of personal profiles, and computation aspects (such as neighborhood formation and prediction generation). One argument why distribution leads to better privacy protection is that users may have better control over their own data if they are stored at the client side as compared to a central (user modeling) server. What is more important though is that CF computation is performed in a distributed and cooperative fashion rather than centrally. Personalization either takes places at the client side using merely the user's data, or is realized by specific privacy-preserving collaborative filtering schemes such as the ones described below.

Yenta (Foner, 1997) is a multi-agent distributed matchmaking system that learns about users by finding sets of keywords that characterize a user's interests. It matches users with similar interests by comparing their keywords without disclosing their identities. If a match is found, the Yenta clients can discretely negotiate to decide whether the matched users would like to reveal their identities to each other. Yenta utilizes anonymity/pseudonymity and encryption in protecting users' privacy.

Olsson (Olsson, 1998) describes a decentralized social filtering model that is built on interactions between collaborative software agents performing content-based filtering. This system is similar to Yenta but differs in its way of measuring similarity between different users via trust rather than interests as in Yenta.

Canny (Canny, 2002a, 2002b) outlined a peer-to-peer collaborative filtering model in which users' profiles are all stored at the client side so that users can fully control their data. The underlying multi-party computation scheme allows a community of users to compute an aggregate of their data (i.e., a singular value decomposition (SVD) model of the user-item matrix) based solely on vector addition so that individual data will not be disclosed. This non-disclosure property is achieved by using techniques including ElGamal encryption, homomorphic encryption and Zero Knowledge Proofs.

Miller et al. (Miller, Konstan, & Riedl, 2004) propose a peer-to-peer CF algorithm called PocketLens. For each individual user, PocketLens first searches for neighbors in the P2P network, then incrementally updates the user's individual item-item similarity model by incorporating one neighbor's ratings at a time (the neighbor's ratings will be discarded after updating the model), and finally generates recommendations based on the model. The paper also compares and discusses five implementation frameworks:

- a central server architecture where the key data is stored on a central server while the computations are performed at each individual node;
- a random discovery architecture that allows users to remain anonymous and uses Gnutella's ping/pong mechanism for finding neighbors;
- a transitive traversal architecture that allows clients to share their neighborhood lists by query flooding and thus enables neighborhood formation via a form of transitivity;
- a content-addressable architecture that adopts P2P file sharing networks, e.g., Chord, which places a deterministic overlay routing system over the network and provides a scalable and distributed lookup function (the II-Chord implementation described in the paper uses the network basically as a distributed storage mechanism to collaboratively build and maintain the item-item matrix); and
- a secure blackboard architecture that leverages the secure operations used in a secure online voting protocol and in Canny's work (Canny, 2002a, 2002b), whereby each client writes encrypted partial results to a Write Once Read Many (WORM) blackboard and the final model is generated by incorporating those partial profiles.

Gilburd et al. (Gilburd, Schuster, & Wolff, 2004) introduce a k-TTP (trusted third party) model which suggests that privacy is preserved as long as no participant of a distributed (joint) computation learns statistics of a group with less than k members. This is less restrictive than an ordinary TTP model in the sense that it does not protect unauthorized access to statistics of individual users if less than k members participate in a joint computation, and is thus more flexible. The authors demonstrate that k-TTP enables more scalable distributed computation schemes. While the paper illustrates the idea of k-TTP by an association-rule mining algorithm, the same idea could be applied to personalization techniques such as collaborative filtering. Berkovsky et al.'s idea of super-peers echoes the same aggregation spirit (Berkovsky, Eytani, Kuflik, & Ricci, 2006).

Privacy-preserving collaborative filtering

The aim of work in this area is to apply and extend privacy-preserving data mining techniques in the area of collaborative filtering. The common approach for achieving privacy preservation in data mining tasks is to replace each message exchange in an ordinary distributed data mining algorithm with a cryptographic primitive that provides the same information without disclosing the data of the individual participants. The research challenge here is to enable users to contribute their information for CF purposes without compromising their privacy (e.g., through exposure of their personal data). Here, privacy-preserving CF is treated as a secure multiparty computation problem where users and different websites jointly conduct CF computations based on their private data. These parties could be mutually untrusted, or even competitors. Typical ways of privacy preservation include decentralization, encryption, aggregation, perturbation and obfuscation.

Encryption

In this type of work, CF computation is based on encrypted user data. An example is the abovementioned work of (Canny, 2002a), which describes a secure multi-party computation scheme that allows a community of users to compute an aggregate of their data without disclosing individual

data by using homomorphic encryption and ElGamal encryption. More specifically, a combination of ElGamal encryption and homomorphic encryption allows vectors to be added by multiplying the encrypted addends, and the final result to be decrypted. Individual addends can be verified as valid data using zero knowledge proofs. The resultant aggregate SVD model can then be used to generate personalization.

Randomized perturbation

Polat and Du (Polat & Du, 2003, 2005a, 2005b) demonstrate the usage of randomized perturbation techniques (adding random numbers from a given range to the original data) in disguising the original user ratings before feeding them into CF algorithms based on correlation and singular value decomposition. The CF system thereby does not know the exact values of the original ratings, yet is still able to compute reasonably accurate recommendations. The underlying reason is that the CF algorithms often use aggregations like scalar products and sums, and that the perturbations tend to cancel themselves out.

Aggregation

In this privacy-protecting approach (e.g., (Canny, 2002a)), users' personal data are aggregated in such a way that an individual's data cannot be identified.

Community model

In this approach, CF computation (e.g., model generation) is carried out collaboratively by a community of clients. The difference to aggregation techniques is that a community model may not generate an aggregate model and may still reveal individual user's data, e.g., in the II-Chord implementation of PocketLens (Miller, Konstan, & Riedl, 2004). Both aggregate and community model can also be considered as examples of distributed personalization, since they either store personal profiles or perform CF computation in a distributed manner.

Obfuscation

Another way of disguising users' personal data is via obfuscation. Berkovsky et al. (Berkovsky, Eytani, Kuflik, & Ricci, 2005) describe a decentralized CF model in which user profiles are stored at the client side. In this approach, some of the personal data is replaced by some other data (which is either constant or drawn from some distribution). The authors demonstrate that relatively large parts of the user profile can be obfuscated while CF can still generate reasonably accurate recommendations. In their follow-up work (Berkovsky, Eytani, Kuflik, & Ricci, 2006), they propose a decentralized recommendation generation scheme that is based on a hierarchical neighborhood topology. More specifically, users (peers) are organized into groups managed by super-peers. To enhance privacy, the super-peers choose only a random subset of their peers to form the neighborhood of similar users. To protect individual peers' privacy within a peer-group, the obfuscation techniques can be used and also only a subset of peers can be queried.

Scrutable personalization

Kay et al. (Kay, 2006; Kay, Kummerfeld, & Lauder, 2003) suggest putting scrutability into user modeling and personalized systems. By scrutability the authors mean that users can understand and control what goes into their user model, what information from their model is available to different services, and how the model is managed and maintained. Their user modeling system Personis applies three privacy-enhancing mechanisms to control the protection of each unit of personal information ("evidence") in the user model (Kay, Kummerfeld, & Lauder, 2003):

- expiration dates and purging of older evidence,
- compaction, for replacing a set of evidence from a single source with an aggregate, and
- morphing, which replaces an arbitrary collection of evidence.

For controlling the usage of evidences from the user model, Personis allows users to restrict the evidences that are available to applications, and the methods that may generate a user model and

operate on it. Despite the desirability of scrutability from a privacy point of view, its implementation and control is currently very challenging, due to users' lack of understanding of these notions and of effective and efficient user interfaces to support them. Moreover, scrutability may reveal the personalization methods that a website uses, which may pose a problem in application areas in which those are considered to be competitive advantages and therefore confidential (e.g., in online retail websites).

Task-based personalization

Herlocker and Konstan (Herlocker & Konstan, 2001) propose a content-independent task-focused recommendation scheme. The scheme assumes that a traditional recommender system may already possess historical ratings data, and that recommendation is possible with data that pertain to the current session or specific task only (e.g., buying a martial arts DVD) rather than collecting a comprehensive profile of the user across multiple sessions. The system builds an item-item association model based on the legacy ratings, and uses the model to generate recommendations. The privacy improvement is that users do not need to disclose their historical ratings while still being able to receive task-focused recommendations. Cranor (Cranor, 2003) also supports task or session based personalization as a way to reduce privacy risks and make privacy compliance easier. However, the price is that the recommendations are not truly personalized, i.e., all users may receive the same recommendations for the same task.

Tailoring personalization to users' privacy constraints

Wang et al. (Wang, Kobsa, van der Hoek, & White, 2006) propose a user modeling server architecture that encapsulates different user modeling components (UMCs) and, at any point during runtime, ascertains that only those components can be operational that are in compliance with the currently prevailing privacy constraints (including privacy legislation, regulations and users' personal privacy preferences). Moreover, the architecture can also dynamically select the component with the optimal anticipated personalization effects among those that are currently permissible (Kobsa, 2003). Each user has their own tailored instance of the UMC pool, containing only those UMCs that meet the privacy requirements for the respective user (users with identical UMC pool instances share the same instance). An advantage of this approach is its capability to reconfigure the architecture immediately to cater to users' changes of privacy preferences at any time (we denote this capability as dynamism support). This approach directly addresses the principles of enforcement, ease of compliance and responsiveness.

Analysis of technical solutions for privacy-enhanced personalization

We have seen that different privacy enhancing solutions for personalized systems often implement several basic techniques. Table 5 gives a summary of the techniques used in the discussed systems. Table 6 shows how well a set of representative privacy protection solutions from the ones discussed above meet the privacy principles described earlier. Table 7 presents how these solutions address the privacy concerns in web personalization described earlier. The following observations can be made:

First, several solutions aim for a balance between privacy and personalization. Examples include pseudonymous personalization, scrutable personalization and dynamic personalization. They all address a handful of privacy concerns and achieve at least reasonably good personalization.

Second, none of the solutions in Table 5 uses all available privacy-enhancing techniques. We believe more comprehensive future solutions will need to incorporate a variety of basic privacy enhancing techniques.

Third, none of the solutions in Table 7 addresses all privacy concerns, except Personis which relies on a "user empowerment" strategy. However, Personis does not address all the concerns effectively. For example, it does not provide comprehensible and effective user interfaces even though most users do not possess mental models of the operation of user modeling systems.

Finally, we find that principles such as onward transfer, enforcement, user preference, negotiation, ease of compliance and responsiveness are currently insufficiently observed. Taking "onward transfer"

as an example, no current privacy-enhancing solution in web personalization allows ”sticky” privacy policies that travel with data so that, e.g., user data cannot be copied and transferred by an entity that is only allowed to read the data. Techniques used in Digital Rights Management (DRM) (Rosenblatt, Trippe, & Mooney, 2001) may be adapted for this purpose.

Table 5. Basic privacy protection techniques used in privacy-enhanced personalization solutions

Technique / **System**	A/P	En	SD	CD	Ag	CM	Pe	Ob	ScS	TP	DS
Yenta	X	X	X	X							
Trust-based Social Filtering (Olsson 1998)			X	X							
PocketLens Central Server				X							
PocketLens Random Discovery	X		X	X							
PocketLens Transitive Traversal	X		X	X							
PocketLens II-Chord			X	X		X					
PocketLens Secure Blackboard		X	X	X	X	X					
k-TTP		X	X	X	X	X					
Privacy Preserving CF (Canny 2002a)		X	X	X	X	X					
Factor Analysis-CF (FA-CF) (Canny 2002b)		X	X	X	X	X					
Random Perturbation-CF							X				
Privacy Enhancing CF (Berkovsky et al. 2005)			X	X				X			
Hierarchical Neighborhood Topology-CF (HNT-CF) (Berkovsky et al. 2006)			X	X	X			X			
Personis			X	X	X		X		X		
Task-based Personalization										X	
Privacy-Tailored Personalization											X

A/P: Anonymity/pseudonymity
En: Encryption
SD: Storage distribution
CD: Computation distribution
Ag: Aggregation
CM: Community model
Pe: Perturbation
Ob: Obfuscation
ScS: Scrutability support
TP: Task-based personalization
DS: Dynamism support

Table 6. An Analysis of Privacy Protection Solutions in Web Personalization

Principle \ Solution	Pseudonymous UMS	Yenta	PocketLens + II-Chord	Canny's FA-CF	HNT-CF	Task-based CF	Personis	Privacy-tailored personaliztion
GENERAL								
Notice/Awareness							++	
Choice/Consent		+	+	+	+		+	+
Enforcement/Redress	+	+		+			+	++
User preference							+	++
Negotiation		+						
Ease of adoption	-			-				+
Ease of compliance								++
Usability							-	
Responsiveness								++
Personalization quality	++	+	++	+	++	+	+	++
IDENTITY								
Anonymity	++		+					
Pseudonymity	++	+		++	+		+	
Unobservability	++		+	+	+			
Unlinkability			+	++				
Deniability					+			
SECLUSION								
Seclusion								
DATA								
Minimization				+	+	++		++
Purpose specification							+	+
Collection limitation							+	
Use limitation	+						+	++
Onward transfer								
Access/Participation							++	
Integrity/accuracy							+	
Security	+	+		+	+			

++: Strong support +: Support -: Negative impact

Table 7. How existing solutions address privacy concerns in web personalization

	Control over data							Seclusion	Protection of identity
	Improper acquisition			Improper use			Improper storage	Unwanted solicitation	Identity fraud/theft
	Improper access	Improper collection	Improper monitoring	Improper analysis	Improper merge	Improper transfer			
Pseudonymous UMS		++	++					+	++
Yenta	+	++	+	+	+	+	+		++
PocketLens + II-Chord	+	++	+	+	+	+	+		
Canny's FA-CF	+	+	+	+	+	+	+		++
HNT-CF	+	+	+	+	+	+	+		++
Task-based Personalization		+	+	+	+	+			+
Personis	++	++	++	++	++	++	++	+	+
Privacy-tailored personalization				++	++		++		+

++: Effective +: Partially effective

DISCUSSION

We discuss the major findings of our survey from two points of views, namely the one of users and of websites.

Users

User would like to enjoy personalized services of websites while at the same time have their individual privacy needs respected (Kobsa, 2007a). The traditional strategy for addressing users' privacy needs is through expression and enforcement – users specify their privacy needs which are then translated into formal expressions and finally enforced in technical solutions.

There are several problems with this strategy. First, privacy decisions (e.g., whether to disclose one's telephone number in a particular situation) are inherently contingent and situated. As Dourish and colleagues (DiGioia & Dourish, 2005; Dourish & Anderson, 2006) point out, the artificial separation of configuration and action may be overly rigid or ineffective. Second, it is a known fact that users' actual behaviors may diverge from their stated privacy attitudes or preferences (Spiekermann, Grossklags, & Berendt, 2001). Third, we observe that currently available technical privacy languages fall short of expressing users' highly flexible and nuanced privacy needs. This may well be an inevitable "social-technical gap" (Ackerman, 2000) between human activities/decisions and what we can support technically. Forth, even if users' privacy decisions could be accurately translated into enforceable specifications, we notice that the majority of existing solutions lack enforcement mechanisms that respond to users' unpredictable changes of privacy decisions in an effective manner.

We see three emerging ways of alleviating or solving these problems:

1. by empowering users to make informed decisions (e.g., by giving them insights into the consequences of their actions through visualizations of system states and events, by enabling them to carry out their privacy decisions rather than merely expressing them through integration of configuration and action (de Paula et al., 2005), or by providing scrutability support in user models (Kay, 2006));
2. by supporting the negotiation between users and websites to reach a consensus on the privacy practices of websites (e.g., (Buffett, Jia, Liu, Spencer, & Wang, 2004; Preibusch, 2006)); and
3. by enabling run-time system variability (Wang, Kobsa, van der Hoek, & White, 2006) as a way to address the responsiveness principle that directly relates to the enforcement problem.

Websites

One of the pressing challenges that websites face today is the need to provide competitive value-added personalized services to its users while complying with a growing number of regulatory privacy requirements. From our survey, we recognize deficiencies in the area of compliance (see Table 7). More specifically, we witness that compliance-related principles such as enforcement and ease of compliance are mostly not addressed, with the exception of a few solutions based on the abovementioned "expression and enforcement" strategy such as in the IBM Tivoli privacy manager (IBM, 2003). From the previous section we can infer though that this approach may run into problems when users become involved.

In the light of this, we coarsely categorize regulatory privacy requirements into two types. The first type consists of requirements that can be met without user involvement (we call them "website-exclusive" requirements). An instance of this type is "*usage data must be erased immediately after each session*" *(*except for very limited purposes) (DE-TML, 2007). The second type consists of requirements that may include privacy decisions of the user (we call them "user-involving" requirements). Examples are "*users must be able to withdraw their consent to the processing of traffic and location data at any time* (EU, 2002)", and "*value-added (e.g. personalized) services based on traffic or location data require the anonymization of such data or the user's consent* (EU, 2002)".

Since "user-involving" requirements can only be fulfilled by users' involvement (giving their consent), we believe that this type of privacy requirements might also be well addressed by using some

of the alternatives to the expression and enforcement approach that were discussed in the previous section. We expect new solutions to emerge in the future that follow these alternate directions.

In contrast, the traditional strategy of expression and enforcement is by and large appropriate and effective for fulfilling the website-exclusive obligations. First, because of its website-exclusiveness, the user empowerment alternative is obviously irrelevant. Second, the separation of expression and enforcement is no longer a problem here, for three reasons: (1) website-exclusive requirements are usually unambiguous and rigid, and thus amenable to accurate formal expressions; (2) there are tools available that can automatically translate textual requirements into specifications in formal languages like P3P (e.g., IBM's Sparcle (Karat, Karat, Brodie, & Feng, 2005)); and (3) once put into effect, privacy laws and regulations are fairly stable, and changes are normally known a few months before they become effective.

While expression can become much easier with support through tools like Sparcle, enforcement is still quite challenging, for the following reasons.

- An effective enforcement mechanism needs to cover the whole lifecycle of user data from collection to usage to transfer, etc.
- In centralized user modeling systems (which collect and supply user information from and to different websites for usually different purposes), the complexities of defining different permissible purposes for collecting and using personal data must be addressed. What is more, since privacy laws can also affect the permissibility of personalization methods used to process user data, the enforcement may involve substituting methods in the user modeling systems at runtime (Wang, Kobsa, van der Hoek, & White, 2006).
- For legacy systems it is likely that privacy had been disregarded during their design and implementation. As with usability, research has revealed though that privacy and security cannot be an afterthought in system design (de Paula et al., 2005; Dourish & Anderson, 2006; Dourish, Grinter, Dalal, Flor, & Joseph, 2004). The support of the enforcement of privacy in legacy systems is therefore likely to be very hard.

SUMMARY AND FUTURE RESEARCH DIRECTIONS

Privacy and web personalization are in tension with each other. The more user data websites collect and utilize, the better are generally the personalized services they provide but the more potential privacy concerns may arise. With the enactment of privacy legislation and regulations worldwide, the conflict is even more acute because personalized websites are obliged to comply with their provisions, which often have remarkable impacts on how personalization may be performed.

In analyzing technical solutions for privacy-enhancing personalization, we propose and apply a multi-faceted approach, consisting of privacy guidelines, privacy concerns, and privacy-enhancing characteristics of these solutions. We relate these facets to each other and reveal trends and identify deficiencies.

Based on our study of existing privacy-enhancing personalization solutions, we suggest the following directions for future research:

- We advocate more recognition of the importance of privacy in web personalization research and practice, and argue that privacy needs be treated as first-class design requirements since (1) regulatory privacy requirements and users' privacy concerns have significant impacts on personalization and its possible benefits, and (2) privacy, like security and usability, is extremely difficult if not impossible to achieve after a system has already been built. Therefore, privacy should be taken into serious consideration from the early onsets of the development process.
- Further research is needed to improve the expression and enforcement approach. With regard to the expression of privacy constraints, two things are desirable. First, a formal language is needed that can sufficiently express potential privacy constraints. As discussed in (Wang & Kobsa, forthcoming), XACML (OASIS, 2005) seems to come close to this vision. However, further studies need to confirm this or/and uncover deficiencies. Secondly, potential privacy constraints should be captured and expressed as they arise, preferably in real time. Users' privacy concerns usually emerge as they interact with a web-based personalized system. Designers of privacy enhanced web personalization should not assume that users can and would express their privacy concern in a

formal privacy language. A hybrid approach of "user empowerment" and "expression and enforcement" might be promising in which users become empowered to act on their contingent privacy needs and possibly also express them in a user-friendly fashion (e.g., in natural language). Thereafter, the system would compile this information into formal expressions that can be executed and enforced. Systematic enforcement is also largely neglected in privacy enhancement in web personalization. Solutions like the IBM Tivoli Privacy Manager need to be adopted.

- While compliance has long been technically framed and treated as a server-side problem, solutions that follow the user empowerment strategy (such as Personis) bear great potential. How to appropriately empower users in the context of web personalization is still an open question, e.g. in light of the fact the users may not be technically savvy. Techniques such as visualization may be useful in this regard.
- Users' privacy needs have been studied predominately in the domain of E-commerce. However, web personalization can also take place in, e.g., E-learning or Ubiquitous Computing, and research is needed to uncover users' privacy needs in these domains as well. Besides, since users' privacy needs and preferences are inherently dynamic and contingent, users' *individual* privacy needs must be taken into account. Solutions that allow for tailored privacy in personalization at runtime seem promising in this regard (Wang & Kobsa, 2007).
- Another promising future direction is usable personal privacy management tools that can help users manage and keep track of the disclosure and usage of their personal information (e.g., by indicating which organization knows what about the user and employs this information for what purposes).

REFERENCES

Ackerman, M. S. (2000). The Intellectual Challenge of CSCW: The Gap between Social Requirements and Technical Feasibility. Human-Computer Interaction, 15 (2), 179-203.

APEC-FIP. (2004). APEC Privacy Framework (No. 2004/AMM/014rev1): Asia-Pacific Economic Cooperation. from http://www.apecsec.org.sg/.

Arlein, R. M., Jai, B., Jakobsson, M., Monrose, F., & Reiter, M. K. (2000). Privacy-Preserving Global Customization. In proceedings of the 2nd ACM Conference on Electronic Commerce (pp. 176-184). Minneapolis, MN.

Berkovsky, S., Eytani, Y., Kuflik, T., & Ricci, F. (2005). Privacy-Enhanced Collaborative Filtering. In proceedings of the PEP05, UM05 Workshop on Privacy-Enhanced Personalization (pp. 75-84). Edinburgh, UK.

Berkovsky, S., Eytani, Y., Kuflik, T., & Ricci, F. (2006). Hierarchical Neighborhood Topology for Privacy-Enhanced Collaborative Filtering In proceedings of the PEP06, CHI06 Workshop on Privacy-Enhanced Personalization (pp. 6-13). Montreal, Canada.

Buffett, S., Jia, K., Liu, S., Spencer, B., & Wang, F. (2004). Negotiating Exchanges of P3P-Labeled Information for Compensation. Computational Intelligence, 20 (4), 663-677.

Camp, J., & Osorio, C. (2003). Privacy-Enhancing Technologies for Internet Commerce. In O. Petrovic, M. Ksela, M. Fallenböck & C. Kittl (Eds.), Trust in the Network Economy (pp. 317-331). Wien: Springer.

Canny, J. (2002a). Collaborative Filtering with Privacy. In proceedings of the IEEE Conference on Security and Privacy (pp. 45- 57). Oakland, CA.

Canny, J. (2002b). Collaborative Filtering with Privacy via Factor Analysis. In proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (SIGIR 2002) (pp. 238 - 245). Tampere, Finland.

Cranor, L. F. (2003). 'I Didn't Buy it for Myself': Privacy and Ecommerce Personalization. In proceedings of the 2003 ACM Workshop on Privacy in the Electronic Society. Washington, DC.

de Paula, R. r., Ding, X., Dourish, P., Nies, K., Pillet, B., Redmiles, D. F., et al. (2005). In the Eye of the Beholder: A Visualization-based Approach to Information System Security. International Journal of Human-Computer Studies, 63 (1-2), 5-24.

DE-TML. (2007). German Telemedia Law. from http://www.gesetze-im-internet.de/tmg/.

DiGioia, P., & Dourish, P. (2005). Social Navigation as a Model for Usable Security. In proceedings of the Symposium on Usable Privacy and Security SOUPS 2005 (pp. 101-108). Pittsburgh, PA.

Dourish, P., & Anderson, K. (2006). Collective Information Practice: Exploring Privacy and Security as Social and Cultural Phenomena. Human-Computer Interaction, 21 (3), 319-342.

Dourish, P., Grinter, R. E., Dalal, B., Flor, J. D. d. l., & Joseph, M. (2004). Security Day-to-Day: User Strategies for Managing Security as an Everyday, Practical Problem. Personal Ubiquitous Computing, 8 (6), 391-401.

EU. (1995). Directive 95/46/EC of the European Parliament and of the Council of 24 October 1995 on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of such Data. Official Journal of the European Communities (23 November 1995 No L. 281), 31ff. from http://158.169.50.95:10080/legal/en/dataprot/directiv/directiv.html.

EU. (2002). Directive 2002/58/EC of the European Parliament and of the Council Concerning the Processing of Personal Data and the Protection of Privacy in the Electronic Communications Sector. from http://register.consilium.eu.int/pdf/en/02/st03/03636en2.pdf.

Foner, L. N. (1997). Yenta: A Multi-Agent Referral-Based Matchmaking System. In proceedings of the International Conference on Autonomous Agents (pp. 301-307). Marina del Rey, CA.

Frankowski, D., Cosley, D., Sen, S., Terveen, L., & Riedl, J. (2006). You Are What You Say: Privacy Risks of Public Mentions. In proceedings of the 29th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval (pp. 565-572). Seattle, WA.

FTC. (2000a). Online Profiling: A Report to Congress. from http://www.ftc.gov/os/2000/07/onlineprofiling.pdf.

FTC. (2000b, May 2000). Privacy Online: Fair Information Practices in the Electronic Marketplace. A Report to Congress. from http://www.ftc.gov/reports/privacy2000/privacy2000.pdf.

FTC. (2000c). The Seven Safe Harbor Principles. from http://europa.eu.int/comm/internal_market/en/media/dataprot/news/shprinciples.pdf.

Gabber, E., Gibbons, P. B., Matias, Y., & Mayer, A. (1997). How to Make Personalized Web Browsing Simple, Secure, and Anonymous. In Financial Cryptography'97 (Vol. 1318). Berlin - Heidelberg - New York: Springer Verlag.

Gilburd, B., Schuster, A., & Wolff, R. (2004). k-TPP: A New Privacy Model for Large-Scale Distributed Environments. In proceedings of the 10th ACM SIGKDD International Conference on Knowledge Discovery and Date Mining (KDD'04) (pp. 563-568). Seattle, WA.

Herlocker, J., & Konstan, J. (2001). Content-Independent Task-Focused Recommendation. IEEE Internet Computing, 5 (6), 40 - 47.

Hitchens, M., Kay, J., Kummerfeld, B., & Brar, A. (2005). Secure Identity Management for Pseudo-Anonymous Service Access. In proceedings of the Security in Pervasive Computing: Second International Conference (pp. 48-55). Boppard, Germany.

IBM. (2003). IBM Tivoli Privacy Manager for E-Business. from http://www-306.ibm.com/software/tivoli/products/privacy-mgr-e-bus/.

Ishitani, L., Almeida, V., & Wagner, M., Jr. (2003). Masks: Bringing Anonymity and Personalization Together. IEEE Security & Privacy Magazine, 1 (3), 18-23. from DOI 10.1109/MSECP.2003.1203218.

Karat, J., Karat, C.-M., Brodie, C., & Feng, J. (2005). Privacy in Information Technology: Designing to Enable Privacy Policy Management in Organizations. International journal of Human-Computer Studies, 63, 153-174. from DOI 10.1016/j.ijhcs.2005.04.011.

Kay, J. (2006). Scrutable Adaptation: Because We Can and Must. In Adaptive Hypermedia and Adaptive Web-Based Systems (pp. 11-19): Springer Berlin / Heidelberg.

Kay, J., Kummerfeld, B., & Lauder, P. (2003). Managing private user models and shared personas. In proceedings of the Workshop on User Modelling for Ubiquitous Computing, 9th International Conference on User Modeling (pp. 1-11).

Kobsa, A. (2002). Personalization and International Privacy. Communications of the ACM, 45 (5), 64-67. from DOI 10.1145/767193.767196.

Kobsa, A. (2003). A Component Architecture for Dynamically Managing Privacy in Personalized Web-based Systems. In proceedings of the Privacy Enhancing Technologies: Third International Workshop (Vol. LNCS 2760, pp. 177-188). Dresden, Germany. from DOI: 10.1007/b94512.

Kobsa, A. (2007a). Privacy-Enhanced Personalization (cover article). Communications of the ACM, 50 (8), 24-33. from DOI 10.1145/1278201.1278202.

Kobsa, A. (2007b). Privacy-Enhanced Web Personalization. In P. Brusilovsky, A. Kobsa & W. Nejdl (Eds.), The Adaptive Web: Methods and Strategies of Web Personalization (pp. 628-670): Springer-Verlag. from DOI 10.1007/978-3-540-72079-9_21.

Kobsa, A., J. Koenemann and W. Pohl. (2001). Personalized Hypermedia Presentation Techniques for Improving Online Customer Relationships. The Knowledge Engineering Review, 16 (2), 111-155. from DOI 10.1017/S0269888901000108.

Kobsa, A., & Schreck, J. (2003). Privacy through Pseudonymity in User-Adaptive Systems. ACM Transactions on Internet Technology, 3 (2), 149-183. from DOI 10.1145/767193.767196.

Malin, B., Sweeney, L., & Newton, E. (2003). Trail Re-Identification: Learning Who You Are From Where You Have Been (Technical Report No. LIDAP-WP12). Pittsburgh, PA: Carnegie Mellon University, Laboratory for International Data Privacy. from http://privacy.cs.cmu.edu/people/sweeney/trails1.pdf.

Miller, B., Konstan, J., & Riedl, J. (2004). PockLens: Toward a Personal Recommender System. ACM Transactions on Information Systems, 22 (3), 437-476.

Nakashima, E. (2006). AOL Search Queries Open Window Onto Users' Worlds. from http://www.washingtonpost.com/wp-dyn/content/article/2006/08/16/AR2006081601751_pf.html.

OASIS. (2005). eXtensible Access Control Markup Language (XACML), Version 2.0; OASIS Standard, February 1, 2005. .

OECD. (1980). Recommendation of the Council Concerning Guidelines Governing the Protection of Privacy and Transborder Flows of Personal Data. from http://www.oecd.org/dsti/sti/it/secur/prod/PRIV-EN.HTM.

Olsson, T. (1998). Decentralised Social Filtering based on Trust. In proceedings of the The AAAI-98 Recommender Systems Workshop (Working Notes) (pp. 84-88). Madison, Wisconsin.

Polat, H., & Du, W. (2003). Privacy-Preserving Collaborative Filtering Using Randomized Perturbation Techniques. In proceedings of the Third IEEE International Conference on Data Mining (ICDM'03) (pp. 625-628). Melbourne, FL.

Polat, H., & Du, W. (2005a). Privacy-Preserving Collaborative Filtering. The International Journal of Electronic Commerce (IJEC), 9 (4), 9-35.

Polat, H., & Du, W. (2005b). SVD-based Collaborative Filtering with Privacy. In proceedings of the 20th ACM Symposium on Applied Computing (pp. 791-795). Santa Fe, NM.

Preibusch, S. (2006). Personalized Services with Negotiable Privacy Policies. In proceedings of the PEP06, CHI 2006 Workshop on Privacy-Enhanced Personalization (pp. 29-38). Montreal, Canada.

Rao, J. R., & Rohatgi, P. (2000). Can Pseudonymity Really Guarantee Privacy? In proceedings of the 9th USENIX Security Symposium (pp. 85–96).

Rosenblatt, W., Trippe, W., & Mooney, S. (2001). Digital Rights Management: Business and Technology. Indianapolis, IN: Hungry Minds, Inc.

Solove, D. (2006). A Taxonomy of Privacy. University of Pennsylvania Law Review, 154 (3), 477-564.

Spiekermann, S., Grossklags, J., & Berendt, B. (2001). Stated Privacy Preferences versus Actual Behaviour in EC Environments: a Reality Check. In proceedings of the WI-IF 2001: the 5th International Conference Wirtschaftsinformatik - 3rd Conference Information Systems in Finance (pp. 129-148). Augsburg, Germany.

Sweeney, L. (2002). k-Anonymity: A Model for Protecting Privacy. International Journal on Uncertainty, Fuzziness, and Knowledge-based Systems, 10 (5), 557-570. from DOI 10.1142/S0218488502001648.

Teltzrow, M., & Kobsa, A. (2004). Impacts of User Privacy Preferences on Personalized Systems: a Comparative Study. In C.-M. Karat, J. Blom & J. Karat (Eds.), Designing Personalized User Experiences for eCommerce (pp. 315-332). Dordrecht, Netherlands: Kluwer Academic Publishers. from DOI 10.1007/1-4020-2148-8_17.

USACM. (2006). USACM Policy Recommendations on Privacy. New York, NY: U.S. Public Policy Committee of the Association for Computing Machinery. from http://www.acm.org/usacm/Issues/Privacy.htm.

Wang, H., Lee, M. K. O., & Wang, C. (1998). Consumer Privacy Concerns about Internet Marketing. Communications of the ACM, 41 (3), 63-70.

Wang, Y., & Kobsa, A. (2006). Impacts of Privacy Laws and Regulations on Personalized Systems. In proceedings of the PEP06, CHI06 Workshop on Privacy-Enhanced Personalization (pp. 44-46). Montréal, Canada. http://www.ics.uci.edu/~kobsa/papers/2006-PEP-wang-kobsa.pdf

Wang, Y., & Kobsa, A. (2007). Respecting Users' Individual Privacy Constraints in Web Personalization In proceedings of the UM07, 11th International Conference on User Modeling (pp. 157–166). Corfu, Greece. from DOI 10.1007/978-3-540-73078-1.

Wang, Y., & Kobsa, A. (forthcoming). Privacy-Enhancing Technologies. In M. Gupta & R. Sharman (Eds.), Social and Organizational Liabilities in Information Security: IGI Global.

Wang, Y., Kobsa, A., van der Hoek, A., & White, J. (2006). PLA-based Runtime Dynamism in Support of Privacy-Enhanced Web Personalization. In proceedings of the 10th International Software Product Line Conference (pp. 151-162). Baltimore, MD. from DOI 0.1109/SPLINE.2006.1691587.

Wang, Y., Zhaoqi, C., & Kobsa, A. (2006). A Collection and Systematization of International Privacy Laws, with Special Consideration of Internationally Operating Personalized Websites. from http://www.ics.uci.edu/~kobsa/privacy/intlprivlawsurvey.html.

ADDITIONAL READINGS

Agrawal, D., & Aggarwal, C. C. (2001). *On the Design and Quantification of Privacy Preserving Data Mining Algorithms*. In proceedings of the 20th ACM SIGACT-SIGMOD-SIGART Symposium on Principles of Database System (pp. 247–255). Santa Barbara, CA.

Agrawal, R., & Srikant, R. (2000). *Privacy-Preserving Data Mining*. In proceedings of the ACM-SIGMOD 2000 Conference on Management of Data (pp. 439-450). Dallas, TX.

Ashrafi, N., & Kuilboer, J.-P. (2005). Privacy Protection via Technology: Platform for Privacy Preferences (P3P). *International Journal of E-Business Research, 1*(2), 56-69. from http://www.igi-online.com/details.asp?ID=7675.

Blarkom, G. W. v., Borking, J. J., & Verhaar, P. (2003). PET. In G. W. v. Blarkom, J. J. Borking & J. G. E. Olk (Eds.), *Handbook of Privacy and Privacy-Enhancing Technologies: The Case of Intelligent Software Agents* (pp. 33-53): College bescherming persoonsgegevens.

Borking, J., & Raab, C. (2001). Laws, PETs and Other Technologies for Privacy Protection. *Journal of Information, Law and Technology*(1). from http://www2.warwick.ac.uk/fac/soc/law/elj/jilt/2001_1/borking/.

Boyens, C., & Günther, O. (2002). *Trust Is not Enough: Privacy and Security in ASP and Web Service Environments*. In proceedings of the Advances in Databases and Information Systems, 6th East European Conference (Vol. 2435). Bratislava, Slovakia.

Burkert, H. (1997). Privacy-Enhancing Technologies: Typology, Critique, Vision. In P. E. Agre & M. Rotenberg (Eds.), *Technology and Privacy: The New Landscape* (pp. 126-143). Boston, MA: MIT Press.

Cassel, L. N., & Wolz, U. (2001). *Client Side Personalization*. In proceedings of the DELOS Workshop: Personalisation and Recommender Systems in Digital Libraries.

Ceri, S., Dolog, P., Matera, M., & Nejdl, W. (2004). Model-Driven Design of Web Applications with Client-Side Adaptation. In N. Koch, P. Fraternali & MartinWirsing (Eds.), *Web Engineering: 4th International Conference, ICWE 2004* (pp. 201-214). Berlin–Heidelberg: Springer Verlag.

Chellappa, R. K., & Sin, R. (2005). Personalization versus Privacy: An Empirical Examination of the Online Consumer's Dilemma. *Information Technology and Management, 6*(2-3), 181-202. from DOI 10.1007/s10799-005-5879-y.

Coroama, V., & Langheinrich, M. (2006). *Personalized Vehicle Insurance Rates: A Case for Client-Side Personalization in Ubiquitous Computing*. In proceedings of the PEP06, CHI 2006 Workshop on Privacy-Enhanced Personalization (pp. 56-59). Montreal, Canada.

Cranor, L. F., & Reidenberg, J. R. (2002). *Can User Agents Accurately Represent Privacy Notices?* In proceedings of the 30th Research Conference on Communication, Information and Internet Policy. Alexandria, VA.

Dornbach, P., & Németh, Z. (2003). Privacy Enhancing Profile Disclosure. In R. Dingledine & P. Syverson (Eds.), *PET 2002* (Vol. LNCS 2482, pp. 85–98). Heidelberg, Germany: Springer-Verlag.

Foner, L. N. (1999). *Political Artifacts and Personal Privacy: The Yenta Multiagent Distributed Matchmaking System*. Massachusetts Institute of Technology.

Gandon, F. L., & Sadeh, N. M. (2004). Semantic Web Technologies to Reconcile Privacy and Context Awareness. *Web Semantics: Science, Services and Agents on the World Wide Web, 1*, 241–260. from http://dx.doi.org/10.1016/j.websem.2003.07.008.

Garfinkel, S. L., Juels, A., & Pappu, R. (2005). RFID Privacy: an Overview of Problems and Proposed Solutions. *IEEE Security & Privacy Magazine, 3*(3), 34-43. from DOI 10.1109/MSP.2005.78.

Goldberg, I., Wagner, D., & Brewer, E. (1997). *Privacy-Enhancing Technologies for the Internet*. In proceedings of the 42nd IEEE Spring COMPCON. San Jose, CA.

Goldberg, I. A. (2003). Privacy-Enhancing Technologies for the Internet, II: Five Years Later. In R. Dingledine & P. Syverson (Eds.), *Privacy Enhancing Technologies – Second International Workshop, PET 2002* (pp. 1-12). Berlin - Heidelberg: Springer Verlag.

ICPP-SNG. (2003). *Identity Management Systems (IMS): Identification and Comparison Study* (Report 2003-09-07): Independent Centre for Privacy Protection (ICPP) Schleswig-Holstein and Studio Notarile Genghini (SNG). from http://www.datenschutzzentrum.de/idmanage/study/ICPP_SNG_IMS-Study.pdf.

Juels, A. (2006). Technological Approaches to the RFID Privacy Problem. In S. Garfinkel & B. Rosenberg (Eds.), *RFID : Applications, Security, and Privacy* (pp. 329-339). Upper Saddle River, NJ: Addison-Wesley.

Karjoth, G., & Schunter, M. (2002). A Privacy Policy Model for Enterprises: Presentation at the 15th Computer Security Foundations Workshop (CSFW'02), Cape Breton, Nova Scotia, Canada, IBM Research Zurich Research Laboratory.

Kobsa, A., Chellappa, R. K., & Spiekermann, S. (Eds.). (2006). *Proceedings of CHI-2006 Workshop on Privacy-Enhanced Personalization*. Montréal, Canada. from http://www.isr.uci.edu/pep06/papers/Proceedings_PEP06.pdf

Kobsa, A., & Cranor, L. (Eds.). (2005). *Proceedings of the UM05 Workshop on Privacy-Enhanced Personalization*. Edinburgh, Scotland. http://www.isr.uci.edu/pep05/papers/w9-proceedings.pdf

Mulligan, D., & Schwartz, A. (1999). *Your Place or Mine? Privacy Concerns and Solutions for Server and Client-Side Storage of Personal Information*. In proceedings of the Computers, Freedom & Privacy Conference (pp. 81-84).

Ramakrishnan, N., J.Keller, B., Mirza, B. J., Grama, A. Y., & Karypis, G. (2001). Privacy Risks in Recommender Systems. *IEEE Internet Computing* (Nov-Dec.), 54-62.

Reagle, J., & Cranor, L. (1999). The Platform for Privacy Preferences. *Communications of the ACM, 42*(2), 48-55.

Schreck, J. (2003). *Security and Privacy in User Modeling*. Dordrecht, Netherlands: Kluwer Academic Publishers.

Senicar, V., Jerman-Blazic, B., & Klobucar, T. (2003). Privacy-Enhancing Technologies: Approaches and Development. *Computer Standards & Interfaces, 25*(2), 147-158.

Tavani, H., & Moor, J. (2001). Privacy Protection, Control of Information, and Privacy-Enhancing Technologies. *ACM SIGCAS Computers and Society, 31*(1), 6-11.

Volokh, E. (2000). Personalization, Privacy, and the First Amendment. In CEI Staff (Ed.), *The Future of Financial Privacy: Private Choices versus Political Rules*: CEI.

Warren, S., & Brandeis, L. D. (1890). The Right to Privacy. *Harvard Law Review, 4*(5), 193-220. from http://www.lawrence.edu/fast/boardmaw/Privacy_brand_warr2. html.

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