A Privacy-Enhancing Architecture for Databases

Kirsten Wahlstrom

School of Computer and Information Science
University of South Australia, GPO Box 2471, Adelaide 5001, South Australia
kirsten.wahlstrom@unisa.edu.au

Gerald Quirchmayr

Department of Distributed and Multimedia Systems
University of Vienna, Dr.-Karl-Lueger-Ring 1, A - 1010 Wien
and
School of Computer and Information Science
University of South Australia, GPO Box 2471, Adelaide 5001, South Australia
gerald.quirchmayr@univie.ac.at

Innovative approaches to searching for hidden meaning in data have recently emerged. In some cases, however, the support of privacy is compromised and consistent privacy protection remains elusive. Privacy perceptions and requirements differ from person to person and over time, leading to diverse privacy preferences within a community. This paper reports on a privacy-enhancing database architecture that is premised upon this view of privacy and on Australia's Use and Disclosure National Privacy Principle. The approach extends to supporting privacy under Knowledge Discovery and Data Mining.

Keywords: Privacy, privacy-aware access control, obligation management, privacy-enhancing technology

ACM classification: K.4.1

1. INTRODUCTION

Knowledge Discovery (KD) and Data Mining (DM) are widely practiced and while providing market advantage (Piatetsky-Shapiro, 2007), can conflict with privacy protection (Køien and Oleshchuk, 2007). Charlesworth (2000) argues

> … private enterprise has increasingly replaced national governments as the largest potential threat to personal data privacy

and many governments have enacted legislation to support privacy (Baumer, Earp and Poindexter, 2004).

Privacy protection has a long history. Discussions of legal approaches for supporting privacy date from the late nineteenth century (Cooley, 1888; Warren and Brandeis, 1890) and are ongoing (Baumer *et al*, 2004). Changes in political conditions and advances in technology continue to provide new challenges (Bouguettaya and Eltoweissy, 2003; Janczewski, 2003; Ohkubo, Suzuki and Kinoshita, 2005) and robust, consistent privacy protection remains elusive (Charlesworth, 2000).

Privacy emerges from a society's communication practices (Westin, 2003). Perceptions of privacy differ from person to person and change according to circumstances (Gavison, 1980), leading to diverse and changing privacy preferences within a community.

Manuscript received: 25 June 2008
Communicating Editor: Ljiljana Brankovic

People maintain their privacy by limiting their accessibility and controlling information that describes them (Solove, 2002; Volokh, 2000). Acquisti and Grossklags (2005) found that users actively control information and inadvertently deteriorate their privacy when information sufficient for reliable decision-making is unavailable.

Although this understanding of privacy is supported by generations of philosophers and legal commentators (Floridi, 2006; Gavison, 1980; Rachels, 1975; Warren and Brandeis, 1890) some areas of privacy research make no distinction between privacy, confidentiality and security. Consistently with the established views outlined above, we consider that natural people require privacy and that organisations require confidentiality and compliance with privacy legislation.

The preference for controlling information (Solove, 2002; Volokh, 2000), the requirement for sufficiently informed decision-making (Acquisti and Grossklags, 2005), the scope for conflict between KD and DM and privacy protection (Køien and Oleshchuk, 2007) and the requirement that organisations comply with privacy legislation suggest an avenue for investigation. This paper reports on the design and development of an approach to supporting privacy that is premised on individual citizens' privacy preferences, Australia's Use and Disclosure National Privacy Principle (NPP2) and privacy constraints required by organisations or obligated by legislation. The approach is a database architecture that extends to datawarehouses and KD and DM.

The next section outlines a survey of Privacy-Enhancing Technologies (PETs) to inform our approach. This is followed by a review of NPP2 to ensure the architecture delivers compliance. The architecture's design is elaborated prior to specification of its rules in predicate logic and then its development is reported. Benefits and limitations conclude the paper.

2. PRIVACY-ENHANCING TECHNOLOGIES

As usees[1] have diverse and changing privacy preferences, data processing can undermine one person's privacy while supporting another's. Thus, privacy protection techniques should be integrated with technologies (Registratiekamer (Netherlands) and Information and Privacy Commissioner (Ontario, Canada), 1995; Taipale, 2004).

PETs aim to support privacy without restricting system functionality (Fischer-Hübner, 2001). This nexus of social concerns and systems design is particularly relevant where usees' informed consent is difficult to ascertain, as in the case of secondary data processing such as KD and DM.

One widely investigated approach addresses DM specifically: Privacy-Preserving Data Mining (PPDM). This section reviews PPDM, anonymity, pseudonymity, profile management, web content filtering and Hippocratic Databases in order to inform, motivate and situate our work.

2.1 Privacy-Preserving Data Mining

PPDM is beneficial in any computational context requiring confidentiality of data, but does not readily extend to the inclusion of privacy preferences of individual citizens. This is because organisations rarely, if ever, have reason to collaborate in computation with individual usees. Furthermore, confidence in KD and DM results depends on sets of data large enough to be viable for statistical analysis and describing a diversity of usee behaviours.

Although a recently emerged research area (Agrawal and Srikant, 2000), PPDM has seen vigorous, diverse research that has motivated survey studies (Verykios *et al*, 2004; Bertino *et al*,

1 Fischer-Hübner (2001) introduces the term 'usee' to represent a person described by data (often referred to as a 'data subject' in other work) whereas a user is the person using a DBMS. This convention is appropriate for a paper in which the distinction between the two groups is relevant and the notion of data being subject to the usee is pursued. Thus, this terminology is adopted throughout.

2005; Wu *et al*, 2007). Verykios *et al* (2004) classify PPDM techniques into three categories: heuristics, cryptography or reconstruction. Bertino *et al* (2005) propose a more detailed taxonomy, which was later extended by Wu *et al* (2007). For simplicity, the taxonomy suggested by Verykios *et al* (2004) is adopted here.

Classification, association rule discovery and clustering are all DM techniques for which heuristic privacy preservation approaches exist. Examples include Atallah *et al* (1999) and Chang and Moskowitz (2000). These approaches exploit the NP-hardness of selective data modification and apply it to DM so that confidentiality is improved.

PPDM approaches using cryptography provide data confidentiality in any DM context in which a risk of exposure is evident. Examples include Clifton *et al* (2002), Ioannidis *et al* (2002) and Lindell and Pinkas (2002).

Finally, reconstruction approaches (examples include Agrawal and Aggarwal, 2001 and Rizvi and Haritsa, 2002) apply perturbation followed by aggregation to support confidentiality. The benefits of all three approaches extend to usees, but consideration of privacy preferences is absent.

2.2 Anonymity and Pseudonymity

Anonymity is provided when a usee's identity cannot be ascertained (Goldberg, Wagner and Brewer, 1997). Anonymity services have been developed for email, web browsing and e-commerce. A disadvantage is that accountability becomes problematic and therefore anonymity services are exploitable by those engaged in criminal activities (Clarke, 2001).

Pseudonymity provides a compromise between anonymity and accountability. Usees employing pseudonyms engage in communications without revealing their identities (Fischer-Hübner, 2001). A trusted third party maintains a cross-reference between usees and pseudonyms and under extenuating circumstances a usee's identity can be retrieved.

2.3 Profile Managers and Web Content Filters

Profiling services disseminate executable web content and cookies to track clickstreams and collect usees' information. These activities are often surreptitious, occurring without the usee's knowledge of the nature of the data being collected, the primary purpose of the collection and any secondary uses to which the data may be put. These services have motivated the development of profile managers and web content filters.

Profile managers enable usees to explicitly control access to data (Cranor *et al*, 2002). A usee sets privacy preferences in their browser once and then the browser shares unrestricted information automatically. However, web servers can respond by not rendering content to a browser that does not share information (Rotenburg, 2001). In this situation usees must compromise their privacy or be prevented from viewing web sites. Effectively, usees are penalised for maintaining stringent privacy preferences.

Browsers also permit usees to opt out of accepting cookies. However, usees' access to information is restricted in those cases where web sites are un-viewable without a cookie being set. Again, usees with stringent privacy preferences are effectively penalised.

Web content filters remove content from web pages which is likely to undermine privacy.

2.4 Hippocratic Databases

Agrawal *et al* (2002) provide a strawman architecture for what they call *Hippocratic Databases*. Their architecture includes usees' privacy metadata, privacy constraints and access controls.

Similarly to Gavison (1980), Solove (2002), Volokh (2000) and Westin (2003), Hodel-Widmer (2006) considers privacy to be apparent in the amount of control a usee has over the data they share with organisations. In making a case for privacy autonomy, Hodel-Widmer implements a Hippocratic Database. The database uses privacy preferences so that usees can regain control of data without hindering the business practices of organisations holding the data.

2.5 Findings

The PETs outlined above support two different privacy conceptualisations. PPDM focuses on confidentiality, whereas anonymity, pseudonymity, profile managers, web content filters and Hippocratic Databases support privacy. None were directly premised on legislation, although they can be used in any context where compliance is sought.

If PPDM is eliminated due to its focus on confidentiality, only Hippocratic Databases suggest extension to KD and DM. Thus, the architecture described in this paper adapts Agrawal *et al*'s (2002) Hippocratic Database concept so that compliance with NPP2 is demonstrated and, similarly to Hodel-Widmer (2006), so that diverse privacy preferences are supported.

3. CONCEPTUALISATION

NPP2 specifies the circumstances under which the use and disclosure of personal information must be regulated. Firstly, use and disclosure are restricted in those circumstances where an organisation has not received prior consent. Secondly, use and disclosure are precluded for sensitive information. These requirements can be met if usees specify consent preferences and disclosure preferences.

If usees' consent preferences are known then data can be organised accordingly. For example, when a usee provides data in an online transaction we can assume that consent for that purpose is granted. However, are they also consenting to secondary, currently unforeseen, uses of their data, such as DM? Organisations may opt to confirm consent in a situation such as this.

Similarly, if usees' disclosure preferences are known then data can be organised accordingly. For example, in the transaction outlined above, a usee can specify whether their data is highly private, private, or not private. Then data can be disclosed to users with correlating access privileges. In addition, if disclosure preferences for data are known, it follows those disclosure preferences for aggregate data, KD and DM findings and the output of other secondary data processing may be calculated.

Clearly, an architecture designed to incorporate usees' privacy preferences is also designed to incorporate privacy requirements of organisations or privacy obligations imposed by legislation. Furthermore, where legacy data are merged with operational data, consent and disclosure preferences for operational data may be analysed and retrospectively applied to legacy data.

Conversely, as new data are propagated to a datawarehouse, new consent and disclosure preferences are also propagated. Consider a situation in which changes in a political climate render sharing of religious affiliation risky. Usees would increasingly consider their religious affiliation highly private, leading to an overall change in the profile of aggregate disclosure preferences. This change can be applied to existing data in the datawarehouse so that any trends in usee preferences that emerge over time can be captured. These changes in privacy preferences can be modelled as constraints that apply to attributes.

The approach proposed here obliges usees to provide consent for data usage and it also obliges them to indicate at data collection which data they consider private and also to what extent these data are private (for example, highly, moderately or not private). While burdening the individual, these obligations facilitate compliance with

- NPP2
- Individual usees' privacy preferences that change over time
- Organisations' privacy preferences and legal obligations

and extend privacy support to datawarehouses and secondary data processing such as DM and KD.

The architecture offers an approach to privacy protection that helps to establish equilibrium between the information requirements of organisations and the privacy requirements of usees. However, if usees' disclosure and consent preferences are omitted, the architecture still provides an effective technique for integrating legal and organisations' privacy requirements with a database and, by extension, with datawarehouses and secondary data processing.

4. DESIGN

The approach suggested here constrains querying and secondary data processing according to the disclosure and consent preferences specified by usees. From this point, precise terminology is used. The phrases *consent and disclosure preferences* or *privacy preferences* will be used in usee contexts and *consent and disclosure constraints* will be used in system contexts.

4.1 Rules

Under the architecture, the privacy preferences specified by usees provide consent and disclosure constraints. Consent constraints enable a subset of data with established consent for unforeseen secondary data processing such as KD and DM. Similarly, disclosure constraints enable subsets of data for disclosure to appropriately privileged users.

The architecture will be realised with rules. There are three sources of rules: from disclosure and consent preferences provided by usees, from organisations' privacy policies and from legal obligations.

In designing the model for consent, secondary data processing is conceptualised as a primitive event that causes rules to be fired. If secondary data processing is executed over an attribute or row with a consent constraint that indicates it is not to be used, then these data may be omitted.

In the design of the disclosure model, users with disclosure privileges issue queries (whether primary or secondary) and data has disclosure constraints. Under the proposed architecture, data will be disclosed to a user if their disclosure privilege is at or above the disclosure constraint of the data. Table 1 provides an overview of this approach.

Of the several DBMS that could be used to implement the architecture, PostgreSQL is selected. It is distributed under the BSD licence and can be extended if necessary. It permits a single data retrieval to be specified as an event, it is fully documented and there is a wide range of support available.

4.2 Privacy constraints

Consent constraints indicate a usee's consent preference for unforeseen secondary data processing. Two possible values exist for this constraint: consent is either provided or not.

Disclosure constraint	Disclosure privilege		
	0	1	2
0	Disclosed	Disclosed	Disclosed
1	Withheld	Disclosed	Disclosed
2	Withheld	Withheld	Disclosed

Table 1: Disclosure of data

Disclosure constraints are modelled on a three-point scale, with zero indicating lower disclosure preferences and three indicating higher disclosure preferences. This narrow range is sufficient for proof-of-concept, however the architecture is readily extended to support a wider and more diverse range of disclosure constraints. For example, legal constraints may be defined over a five-point scale, attribute constraints over a seven-point scale and so on.

A further concern is the scope of disclosure constraint applicability. Usees' preferences may apply to either individual datum or to their row. Apart from further encumbering usees at data collection and additional computational performance overhead, the choice is arbitrary. Therefore, usees' constraints apply to rows.

4.3 Users' Disclosure Privileges

Disclosure privileges will be implemented in PostgreSQL's data dictionary. Users and groups will be created, users will be assigned to groups and then groups will be allocated disclosure privileges. As a three-point scale of disclosure constraints has been specified, three user groups are required.

4.4 Conflict Resolution

As usees and organisations can have competing interests, conflict resolution is required. The goal of the architecture is to minimise the risk to privacy, so data suppression is optimal.

4.5 A Performance Consideration

Computational overhead requires some consideration. If a rule is implemented so that it fires on the retrieval of individual rows, it can be surmised that a considerable performance overhead will arise.

PostgreSQL addresses this performance problem with its lock escalation facility. When data are effected by a rule, they are locked and when the majority of data for an attribute are effected by a rule, the lock is escalated so that it applies to the attribute, rather than to individual data. As a result, this performance issue does not require further consideration here.

5. FORMAL RULE DEFINITIONS

There are four contexts requiring rules.

1. Regulating secondary use with consent constraints.
2. Regulating disclosure with disclosure privileges and
   a. Usees' disclosure constraints
   b. Organisation and legal constraints
   c. Attribute constraints

   For each of these contexts formal rule definitions in predicate logic are provided.

5.1 Consent Constraints

As noted above, consent constraints enable a subset of data for which consent for secondary purposes has been granted. This concept and its inverse are defined by the following predicates:

$$\forall r: \; consent(r) \rightarrow use(r) \tag{1}$$

$$\forall r: \; \neg consent(r) \rightarrow \neg use(r) \tag{2}$$

In these predicates, consent is defined over rows, as outlined above in Section 4.2. The predicates state that for all rows, r, if consent for r exists, then r may be used in secondary data processing. Otherwise, it may not be used.

5.2 Usees' Disclosure Constraints

When a query is executed, disclosure constraints and disclosure privileges determine which data may be disclosed to the user (see Table 1). Consistent with Section 4.2, usee disclosure constraints are defined over rows.

$$\forall r: \forall u: >(DC(r), DP(u)) \rightarrow \neg disclose(r) \tag{3}$$

$$\forall r: \forall u: \neg>(DC(r), DP(u)) \rightarrow disclose(r) \tag{4}$$

These predicates state that for all rows, r, and all users, u, if the disclosure constraint for r is greater than the disclosure privilege for u, then r may not be disclosed, and vice versa.

5.3 Organisation and Legal Constraints

There are three possibilities for specifying predicates in this context: a constraint per row in a table, a constraint per table and a constraint per database. Subsequent levels provide greater breadth and less precision of privacy protection.

While all three options are implementable, legislation is broadly applicable and is not intended for use in detailed contexts. Also, in keeping with the premise that privacy is a requirement of individual people, only usees are qualified to specify disclosure constraints over rows.

It is likely that tables in a database will require a diversity of disclosure constraints. For example, consider a database in a hospital: it may have a table of patients' medical records that refers to a table of medications and their side effects. Clearly the table of medical records requires a much higher disclosure constraint than the table of medications and side effects. However, it is less likely that an entire database would require only one disclosure constraint. Thus, it emerges that organisation and legal disclosure constraints defined over tables are optimal.

However, constraints on tables create implementation challenges. PostgreSQL does not support disclosure constraints defined over a table and located in that table. A system table to maintain a list of operational tables and their disclosure constraints will solve this problem.

Furthermore, as disclosure constraints are to be defined over three values, three tables are required for adequate demonstration of the architecture's viability. Thus, four tables are proposed: three for operational data and one system table for organisation and legal constraints applied to operational tables. Data will be propagated to the operational data tables and each table will have a differing disclosure constraint value.

Similarly to predicates (3) and (4) above, the predicates are as follows.

$$\forall t: \forall u: >(DC(t), DP(u)) \rightarrow \neg disclose(t) \tag{5}$$

$$\forall t: \forall u: \neg>(DC(t), DP(u)) \rightarrow disclose(t) \tag{6}$$

These predicates state that for all tables, t, and all users, u, if the disclosure constraint for t is greater than the disclosure privilege for u, then t may not be disclosed, and vice versa.

5.4 Attribute Constraints

To support privacy preferences that change over time, the architecture includes ongoing re-calculation of attribute constraints. This will enable trends in disclosure constraints on rows to be escalated to attributes.

Clearly, the accuracy of an attribute constraint is dependent upon the range, accuracy and number of disclosure constraints over which it is calculated. That is, greater accuracy in an attribute

constraint can be achieved by providing greater scope for usees to specify exact privacy preferences and from a greater number of disclosure constraints for the attribute.

Similarly to predicates (3) and (4) above, the predicates for specifying the logic are as follows.

$$\forall a: \forall u: >(DC(a), DP(u)) \rightarrow \neg disclose(a) \quad (7)$$

$$\forall a: \forall u: \neg >(DC(a), DP(u)) \rightarrow disclose(a) \quad (8)$$

These predicates state that for all attributes, a, and all users, u, if the disclosure constraint for a is greater than the disclosure privilege for u, then a may not be disclosed, and vice versa.

6. IMPLEMENTATION

As data are disclosed via the select operator, any disclosure rules must be defined to fire on a select. However, PostgreSQL places restrictions on select. Firstly, conditions are not allowed in rules designed to fire on a select: a single, unconditional selection of data is the only action permitted. Furthermore, the unconditional select must reference the same table over which the select operation is defined.

However, in PostgreSQL there is no semantic difference between a select rule and a view definition that allows conditional selects. Views can be defined with conditions creating any subset of data, which means that data can be organised into disclosure levels and consent sets.

The implementation environment was configured, Cygwin and PostgreSQL were installed, PostgreSQL's server was launched and a database called *prototype* was implemented on the server.

The 'psql interactive terminal' was used to create tables, users and user groups, migrate data, and implement rules via the creation of views. The architecture's tables were created, the data were prepared in Excel and saved in a tab-delineated format. These data were then migrated to the operational tables and views for all four rule contexts were constructed over the data. These views were merged to create a master view for each disclosure privilege.

7. DISCUSSION

Thirty-eight tests were run over the architecture to ensure it met design goals. Initially six of the tests for organisation and legal constraints failed; the system was debugged and subsequent tests passed.

Three users were created and each user was granted a different disclosure privilege. The database was populated with 355 rows of data. Each row, attribute or table was assigned a disclosure constraint of 0, 1 or 2. The user with disclosure privilege 0 accessed data with disclosure constraint 0 (177 rows), the user with disclosure privilege 1 accessed data with disclosure constraints 0 and 1 (265 rows) and the user with disclosure privilege 2 accessed all data.

Figures 1 and 2 are screenshots of the data for a user with the lowest and the highest disclosure privileges respectively. Note that the number of rows differs according to disclosure privilege, as do the disclosed attributes and tables.

7.1 Limitations

The disablement of rules under special circumstances was not considered. For example, an extraordinary operational context may require data originally provided by a convicted criminal under a highly private constraint to be disclosed to a user authorised to view only moderately private data. Similarly, abuse of data by authorised users has not been considered.

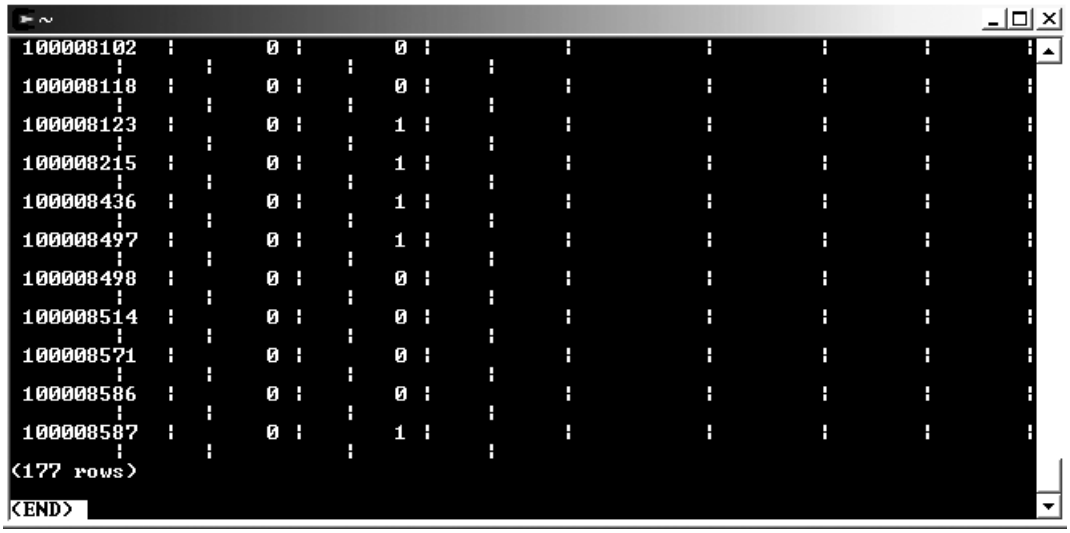

Figure 1: View for disclosure privilege 0

Figure 2: View for disclosure privilege 2

An unavoidable problem is that informed consent becomes less meaningful, as secondary data processing proliferates (Cushman, 1996). Techniques for contending with this limitation restrict the expansion of secondary data processing and therefore fail to achieve equilibrium between organisations and usees.

With respect to consent constraints, there is a risk that data in one of the architecture's views may not be representative of that held in the database, leading to low confidence in secondary data processing.

7.2 Benefits

The architecture supports privacy and is compliant with NPP2. Consent and disclosure constraints facilitate views of data that selectively disclose data to users. The architecture may be readily extended to include disclosure constraints required by any interested party.

Individuals can specify whether and to what extent their data are private, rather than privacy being stipulated by policy. Thus the prototype is consistent with the view that people develop unique, independent and changing perceptions of their privacy.

Enabling individuals to specify privacy preferences leads to the enhancement of an organisation's trustworthiness. Furthermore, organisations may specify privacy preferences that extend those put forth in the legislation, developing trust even further.

The approach described here supplements data with privacy and consent constraints that can be propagated to datawarehouses. Thus, secondary data processing can be constrained, further mitigating the risks to privacy.

8. CONCLUSION

This paper reports on a privacy-enhancing database architecture. The architecture establishes equilibrium between the privacy concerns of individual citizens, regulatory bodies and the data collection interests of business entities.

It is premised on usees' preference for controlling information, the requirement for sufficiently informed decision-making, the scope for conflict between KD and DM and privacy protection and the requirement that organisations comply with privacy legislation. It extends from databases to datawarehouses and secondary data processing. Through the use of attribute constraints, it also allows for privacy preferences that change over time. Future research possibilities include the disablement of rules in response to special circumstances and ensuring the quality of data in the consent set so that secondary data processing is more reliable.

REFERENCES

ACQUISTI, A. and GROSSKLAGS, J. (2005): Privacy and rationality in individual decision-making. *Security & Privacy* 3(1): 26–33.

AGRAWAL, D. and AGGARWAL, C.C. (2001): On the design and quantification of privacy preserving data mining algorithms. *Proceedings of the 20th ACM Symposium on Principles of Database Systems*, Santa Barbara, US, 247–255, ACM Press.

AGRAWAL, R., KIERNAN, J., SRIKANT, R. and XU, Y. (2002): Hippocratic databases. *Proceedings of the 28th International Conference on Very Large Databases*, Hong Kong, China, 143–154, Morgan Kaufmann Publishers.

AGRAWAL, R. and SRIKANT, R. (2000): Privacy-preserving data mining. *SIGMOD Record* 29(2): 439–450.

ATALLAH, M.J., BERTINO, E., ELMAGARMID, A.K., IBRAHIM, M. and VERYKOIS, V.S. (1999): Disclosure limitation of sensitive rules. *Proceedings of the IEEE Knowledge and Data Engineering Workshop*, Chicago, US, 45–52, IEEE Computer Society Press.

BAUMER, D.L., EARP, J.B. and POINDEXTER, J.C. (2004): Internet privacy law: a comparison between the United States and the European Union. *Computers & Security* 23(5): 400–412.

BERTINO, E., FOVINO, I.N. and PROVENZA, L P. (2005): A framework for evaluating privacy preserving data mining algorithms. *Data Mining and Knowledge Discovery* 11(2): 121–154.

BOUGUETTAYA, A. and ELTOWEISSY, M. (2003): Privacy on the web: facts, challenges, and solutions. *Security & Privacy* 1(6): 40–49.

CHANG, L. and MOSKOWITZ, I.S. (2000): An integrated framework for database inference and privacy protection. *Proceedings of Data and Applications Security: Developments and Directions. IFIP TC11/WG11.3 Fourteenth Annual Working Conference on Database Security*, Schoorl, Netherlands, 161–172, Kluwer Aademic.

CHARLESWORTH, A. (2000): Clash of the data titans? US and EU data privacy regulation. *European Public Law* 6(2): 253–274.

CLARKE, R. (2001): Introducing PITs and PETs: technologies affecting privacy. *Privacy Law & Policy Reporter* 7(9): 181–183.

CLIFTON, C., KANTARCIOGLOU, M., LIN, X. and ZHU, M.Y. (2001): Tools for privacy preserving distributed data mining. SIGKDD *Explorations* 4(2): 28-34.

COOLEY, T. (1888): A Treatise on the Law of Torts or the Wrongs which arise independent of contract. 2nd edn, Callaghan.
CRANOR, L., LANGHEINRICH, M., MARCHIORI, M., PRESLER-MARSHALL, M. and REAGLE, J. (2002): The platform for privacy preferences 1.0 (P3P1.0) specification: W3C working draft 28 September 2001. http://www.w3.org/TR/P3P/. Accessed 15 Aug 2008.
CUSHMAN, R. (1996): Information and medical ethics: protecting patient privacy. *IEEE Technology and Society Magazine* 15(3): 32–39.
FISCHER-HÜBNER, S. (2001): IT-security and privacy: design and use of privacy-enhancing security mechanisms, Springer-Verlag.
FLORIDI, L. (2006): Four challenges for a theory of informational privacy. *Ethics and Information Technology* 8(3): 109–119.
GAVISON, R. (1980): Privacy and the limits of law. *Yale Law Journal* 89(3): 421–471.
GOLDBERG, I., WAGNER, D. and BREWER, E. (1997): Privacy-enhancing technologies for the Internet. *Proceedings of the 42nd IEEE International Computer Conference: Hot Systems, Cool Software*, San Jose, US, 103–109, IEEE Computer Society Press.
HODEL-WIDMER, T.B. (2006): Designing databases that enhance people's privacy without hindering organizations. *Ethics and Information Technology* 8(1): 3–15.
IOANNIDIS, I., GRAMA, A. and ATALLAH, M. (2002): A secure protocol for computing dot-products in clustered and distributed environments. *Proceedings of the International Conference on Parallel Processing*, Vancouver, Canada, 379–384, IEEE Computer Society Press.
JANCZEWSKI, L. J. (2003): New challenges in privacy protection. In *Advanced topics in global information management*. 125–139. Tan, F. B. (ed). IGI Publishing.
KØIEN, G. and OLESHCHUK, V. (2007): Personal privacy in a digital world. *Telektronikk* 103(2): 4–19.
LINDELL, Y. and PINKAS, B. (2002): Privacy preserving data mining. *Journal of Cryptology* 15(3): 177–206.
OHKUBO, M., SUZUKI, K. and KINOSHITA, S. (2005): RFID privacy issues and technical challenges. *Communications of the ACM* 48(9): 66–71.
PIATETSKY-SHAPIRO, G. (2007): Data mining and knowledge discovery 1996 to 2005: Overcoming the hype and moving from "university" to "business" and "analytics". *Data Mining and Knowledge Discovery* 15(1): 99–105.
RACHELS, J. (1975): Why privacy is important. *Philosophy and Public Affairs* 4(4): 323–333.
REGISTRATIEKAMER (NETHERLANDS) and INFORMATION AND PRIVACY COMMISSIONER (ONTARIO, CANADA) (1995): Privacy-enhancing technologies: the path to anonymity, vol 1. http://www.ipc.on.ca/. Accessed 19 Feb 2002.
RIZVI, S.J. and HARITSA, J.R. (2002): Maintaining data privacy in association rule mining. *Proceedings of the 28th International Conference on Very Large Databases*, Hong Kong, China, 682–693, Morgan Kaufmann Publishers.
ROTENBURG, M. (2001): Fair information practices and the architecture of privacy (what Larry doesn't get). *Stanford Technical Law Review* 1.
SOLOVE, D. J. (2002): Conceptualising privacy. *California Law Review* 90(4): 1087–1155.
TAIPALE, K. A. (2004): Technology, security and privacy: the fear of Frankenstein, the mythology of privacy and the lessons of King Ludd. *Yale Journal of Law and Technology* 7(123): 123–201.
VERYKIOS, V.S., BERTINO, E., FOVINO, I. N., PROVENZA, L.P., SAYGIN, Y. and THEODORIDIS, Y. (2004): State-of-the-art in privacy preserving data mining. *SIGMOD Record* 33(1): 50–57.
VOLOKH, E. (2000): Personalization and privacy. *Communications of the ACM* 43(8): 84–88.
WARREN, S.D. and BRANDEIS, L.D. (1890): The right to privacy. *Harvard Law Review* 4(5): 193–220.
WESTIN, A.F. (2003): Social and political dimensions of privacy. *Journal of Social Issues* 59(2): 431 – 453.
WU, X., CHU, C.-H., WANG, Y., LIU, F. and YUE, D. (2007): Privacy preserving data mining research: current status and key issues. *Proceedings of the 7th International Conference on Computational Science-ICCS 2007, Part III (Lecture Notes in Computer Science Vol 4489)*, Beijing, China, 762–772, Springer-Verlag.

BIOGRAPHICAL NOTES

Kirsten Wahlstrom is a lecturer in the School of Computer and Information Science at the University of South Australia. She has been involved in developing and teaching a number of courses and she is named in a Carrick Citation, two Chancellor's Commendations for Community Service, a National Woman of the Year nomination and a UniSA teaching citation. Her research interests include informational privacy, computer ethics and the semantic web. She is a member of the Security Laboratory in UniSA's Advanced Computing Research Centre.


Kirsten Wahlstrom


Gerald Quirchmayr

Professor Gerald Quirchmayr holds doctors degrees in computer science and law from Johannes Kepler University in Linz (Austria) and currently is Professor in the Faculty of Computer Science at the University of Vienna (Austria). In 2001/2002 he held a Chair in Computer and Information Systems at the University of South Australia. He first joined the University of Vienna in 1993 from the Institute of Computer Science at Johannes Kepler University in Linz (Austria) where he had previously been teaching. In 1989/1990 he taught at the University of Hamburg (Germany).

His wide international experience ranges from the participation in international teaching and research projects, very often UN- and EU-based, several research stays at universities and research centres in the US and EU Member States to extensive teaching in EU staff exchange programs in the United Kingdom, Sweden, Finland, Germany, Spain, and Greece, as well as teaching stays in the Czech Republic and Poland. International teaching and specialist missions include UN-coordinated activities in Egypt, Russia and the Republic of Korea. He has served as a member of program committees of many international conferences, chaired several of them, has contributed as reviewer to scientific journals and has also served on editorial boards. He is a member of the Austrian and German computer societies and a member of IFIP working groups. For his contributions to the international IT community he received the IFIP Silver Core Award in 1995.

His major research focus is on information systems in business and government with a special interest in applications, formal representations of decision making, legal issues and IT security. His publication record comprises over 100 peer reviewed papers plus several edited books and conference proceedings as well as nationally and internationally published project reports. In July 2002 he was appointed as Adjunct Professor at the School of Computer and Information Science of the University of South Australia. Since January 2005 he heads the Department of Distributed and Multimedia Systems, Faculty of Computer Science, at the University of Vienna.