Introduction to information privacy

1 Course plan

This document describes the course plan, goals and organization of the course Introduction to information privacy.

1.1 Goals

The course provides understanding of the area of privacy. Passing this course means that a student should,

- recognize the privacy terminology
- identify the principles and motivation of the research in the field of privacy
- classify privacy-enhancing technologies and reflect about their fundamentals
- apply privacy-enhancing technologies in different contexts
- recognize main concepts in database privacy
- describe major data protection methods for different types of data;
- apply techniques for measuring disclosure risk and information loss / data utility for protected data

1.2 Prerequisites

MSc in mathematics or computer science/engineering, with some background in Security, and knowledge of maths.

1.3 Contents

The goal of data privacy and privacy enhancing technologies is to develop theories and technologies that permit the access to information avoiding the disclosure of sensitive information. Different approaches focus on different answers to the questions of what sensitive data are, how sensitive data are represented, and how an adversary can recover sensitive data.

Privacy preserving data mining (PPDM) and statistical disclosure control (SDC) are two areas of research that focus on the problems of data privacy.

In this course, the main concepts to be studied are related to data privacy. Major approaches for data protection will be reviewed and there will be discussions on how these methods ensure privacy. Also subjects to discussions are: disclosure risk measures that evaluate to which extent protected data ensures confidentiality, and information loss measures that evaluate whether protected data is still useful for their analysis.

Concrete examples to be discussed are data protection methods for numerical and categorical database protection, and measures for information loss and disclosure risk for these types of databases. We will also discuss some examples on other types of databases as e.g. online social networks.

1.4 Organization

The course consists of two parts: an introduction to privacy and privacy-enhancing technologies and a deeper study in a selected research area in the field of privacy: database privacy. Both parts are focused on lectures, the study of relevant literature, discussions, group activities and practical assignments. The practical part consists of a project work in the area of information privacy, which may, but not necessarily, be related to the content of the lectures. Participants are required to plan and document their work.

Part I: Introduction to Privacy and Privacy-Enhancing Technologies

1. Introduction to Privacy
   - Definition of Privacy
   - Legal Aspects
   - Research Areas in Privacy
2. Anonymity and Pseudonyms
   - Identity and Identifiers
   - Types of Pseudonyms
   - Anonymity
3. Anonymous Communication Mechanisms
   - The Dining Cryptographers
   - MIX Networks
   - Crowds
   - Onion Routing and TOR
4. An Introduction to an ongoing research area in the field of privacy

Part II: Privacy Aspects in Data Mining

1. For those unfamiliar:
   - Elements of machine learning
   - Clustering and comparison of partitions
   - Association rules
2. Data Privacy Dimensions
   - Owner privacy
   - Respondent and owner privacy
   - Data-driven or general-purpose

- Computation-driven or specific-purpose
- Result-driven

3. Data protection methods
4. Information loss measures
5. Disclosure risk measures

Schedule overview (tentative)

Class 1. (Wed Oct, 2nd 10-12) Introduction to Privacy and Privacy-Enhancing Technologies. Introduction to Privacy. Anonymity and Pseudonyms. Anonymous Communication Mechanisms.

Class 2. (Th Oct, 3rd 13-15) Elements of machine learning. Clustering and comparison of partitions. Association rules.

Class 3. (Mon Oct, 7th 9-12) Data privacy dimensions. Owner privacy. Respondent and owner privacy. Data-driven or general-purpose. Computation-driven or specific-purpose. Result-driven.

Class 4. (Tue Oct, 8th 9-12) Disclosure risk measures. Data protection methods (part I).

Class 5. (Th Oct, 10th 9-12) Data protection methods. Information loss measures. Disclosure risk measures revisited.

1.5 Grading

Final grade in this course is based on the final exam (40%) and the homework assignment (60%).

1.6 Contact

Vicenç Torra, vtorra@iiia.csic.es

References

[1] C. C. Aggarwal, P. S. Yu (Eds.) Privacy-Preserving Data Mining: Models and Algorithms, Springer, 2008.

[2] G. T. Duncan, M. Elliot, J. J. Salazar-Gonzalez, Statistical Confidentiality: Principles and Practice, Springer, 2011.

[3] A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, E. S. Nordholt, K. Spicer, P.-P. de Wolf, Statistical Disclosure Control, Wiley, 2012.

[4] A. Pfitzmann, M. Hansen. A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. v0.34, 10 Aug 2010.

[5] S. Fischer-Hbner, D. Kesdogan, L. Martucci. Privacy and privacy-enhancing technologies. In: S. Furnell, S. Katsikas, J. Lopez, A. Patel (Eds.) Securing Information and Communication Systems: Principles, Technologies, and Applications. Artech House, Norwood, MA, USA. ch.11, p.213-242.

[6] P. Syverson, D. Goldschlag, M. Reed. Anonymous connections and onion routing. In Proc. of the 1997 IEEE Symposium on Security and Privacy (S&P 1997), pages 4454. IEEE Computer Society, 4-7 May 1997.

[7] M. Reiter, A. Rubin. Crowds: anonymity for web transactions. ACM Transactions on Information and System Security (TISSEC), 1(1):66-92, 1998. ISSN 1094-9224. doi: http://doi.acm.org/10.1145/290163.290168