Introduction to Information Privacy

Vicenç Torra

October, 2013

Institut d'Investigació en Intel·ligència Artificial (IIIA-CSIC), Bellaterra, Catalonia

Outline

Contact information

Contact information

Contact:

Contact information

Contact:

- E-mail:
  - vtorra@iiia.csic.es
  - vtorra@ieee.org (not so used)

Contact information

Contact:

- E-mail:
  - vtorra@iiia.csic.es
  - vtorra@ieee.org (not so used)
- Web page with the slides:
  - http://www.ppdm.cat/dp

Contact information

Contact:

- E-mail:
  - vtorra@iiia.csic.es
  - vtorra@ieee.org (not so used)
- Web page with the slides:
  - http://www.ppdm.cat/dp
- My web page:
  - http://www.iiia.csic.es/~vtorra

Description of the course

Description of the course

Goals:

- Understanding of the area of data privacy

Description of the course

Goals:

- Understanding of the area of data privacy
  - 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

Description of the course

Contents: Data Privacy

- Description of the field.
- Definitions and different approaches.
- Areas.
  - Security and privacy
  - Privacy preserving data mining (PPDM)
  - Statistical disclosure control (SDC)

Description of the course

Contents: Data Privacy

- The main concepts to be studied are related to data privacy.
  - Major approaches for data protection will be reviewed
  - Disclosure risk measures
    evaluate to which extent protected data ensures confidentiality,
  - Information loss measures
    evaluate whether protected data is still useful for their analysis.
- Concrete examples for numerical and categorical database protection
  - Data protection methods
  - Measures for information loss and disclosure risk
- Examples on other types of databases as e.g. online social networks.

Description of the course

Organization:

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

Description of the course

Organization:

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

Description of the course

Schedule overview:

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.

Description of the course

Grading:

- Final exam 40%
- Homework assignment 60%

Description of the course

Grading:

- Final exam 40%
- Homework assignment 60%

Assignement:

- Downloadable from: http://www.ppdm.cat/dp/
- File: http://www.ppdm.cat/dp/assignment.pdf

Description of the course

Grading:

- Final exam 40%
- Homework assignment 60%

Assignement:

- Downloadable from: http://www.ppdm.cat/dp/
- File: http://www.ppdm.cat/dp/assignment.pdf
- **Exercise.** Select a data file, protect it using at least two data protection methods and different parameterizations, and plot a R-U map for at least 10 pairs (method, parameter) using a particular information loss and disclosure risk measure. Discuss the results.

Description of the course

Bibliography:

- A. Pfitzmann, M. Hansen. A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. v0.34.
  Terminology
- C. C. Aggarwal, P. S. Yu (Eds.) Privacy-Preserving Data Mining: Models and Algorithms, Springer, 2008.
  PPDM: perturbative approaches and only a few cryptographic tools
- J. Vaidya, C. W. Clifton, Y. M. Zhu (2006) Privacy Preserving Data Mining, Springer.
  PPDM: cryptographic tools

Description of the course

Bibliography:

- A. Pfitzmann, M. Hansen. A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. v0.34.
  Terminology
- C. C. Aggarwal, P. S. Yu (Eds.) Privacy-Preserving Data Mining: Models and Algorithms, Springer, 2008.
  PPDM: perturbative approaches and only a few cryptographic tools
- J. Vaidya, C. W. Clifton, Y. M. Zhu (2006) Privacy Preserving Data Mining, Springer.
  PPDM: cryptographic tools
- G. T. Duncan, M. Elliot, J. J. Salazar-Gonzalez, Statistical Confidentiality: Principles and Practice, Springer, 2011.
- A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, E. S. Nordholt, K. Spicer, P.-P. de Wolf, Statistical Disclosure Control, Wiley, 2012.
  data privacy: perturbative approaches for databases

Description of the course

Bibliography:

- S. Fischer-Hübner, 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.
  Privacy-enhancing technologies

Description of the course

Bibliography:

- S. Fischer-Hübner, 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.

  Privacy-enhancing technologies
- 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 44â54. 4-7 May 1997.
- 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.

  Anonymity in communications