Data privacy. Introduction

Vicenç Torra

February, 2018

SAIL + PICS, School of Informatics, University of Skövde, Sweden

Introduction

Goals:

- Master the privacy terminology
- Identify the principles and motivations of the research field of privacy
- Classify privacy-enhancing technologies and reflect about their fundamentals
- Describe major data protection mechanisms for different types of data
- Apply data techniques, measures of disclosure risk and information loss / data utility to a database

Introduction

Contents: (Data privacy)

- Description of the field
- Definitions and different approaches
- Technical perspective
  (i.e., out: laws, social, and psychological issues)
- *Integrated* technical perspective
  - Statistical disclosure control (SDC)
  - Privacy preserving data mining (PPDM)
  - Privacy enhancing technologies (PET) – Communications

Introduction

Contents:

- Motivation
- Terminology
- Classification of protection methods: roadmap
- Privacy models (overview)
- User privacy
- Computation-driven methods:
  Differential privacy, cryptographic approaches
- Privacy models and disclosure risk measures
- Data-driven methods:
  Masking methods
- Information loss measures

References

- Slides and other material here: http://ppdm.cat/dp/
- Also, examples using packages sdcMicro and sdcTable in R for microdata protection and for tabular data protection.

References

- References:

  - V. Torra, Data Privacy: Foundations, New Developments and the Big Data Challenge, Springer, 2017.
  - T. Benschop, C. Machingauta, M. Welch, Statistical disclosure control for microdata: A practical guide, 2016.
  - A. Hundepool, J. Domingo-Ferrer, L. Franconi, S. Giessing, E. S. Nordholt, K. Spicer, P.-P. de Wolf, Statistical Disclosure Control, Wiley, 2012.
  - M. Templ, Statistical disclosure control for microdata: Methods and applications in R, Springer, 2017.
  - A. Pfitzmann, M. Hansen. A terminology for talking about privacy by data minimization: anonymity, unlinkability, undetectability, unobservability, pseudonymity, and identity management. v0.34.
  - C. C. Aggarwal, P. S. Yu (Eds.) Privacy-Preserving Data Mining: Models and Algorithms, Springer, 2008. mainly perturbative approaches
  - J. Vaidya, C. W. Clifton, Y. M. Zhu (2006) Privacy Preserving Data Mining, Springer.
  - J. Castro, Recent advances in optimization techniques for statistical tabular data protection, European Journal of Operational Research 216 (2012) 257-269.

Evaluation

- Evaluation
  - Database protection, trade-off data protection vs. utility (e.g. in R)
  - Reading papers on methods
  - Prepare a document explaining the results, and present it
- Also,
  - Work on developing a masking method or an attack.