Lecture Notes in Computer Science 5262

Commenced Publication in 1973
Founding and Former Series Editors:
Gerhard Goos, Juris Hartmanis, and Jan van Leeuwen

Editorial Board

David Hutchison
Lancaster University, UK

Takeo Kanade
Carnegie Mellon University, Pittsburgh, PA, USA

Josef Kittler
University of Surrey, Guildford, UK

Jon M. Kleinberg
Cornell University, Ithaca, NY, USA

Alfred Kobsa
University of California, Irvine, CA, USA

Friedemann Mattern
ETH Zurich, Switzerland

John C. Mitchell
Stanford University, CA, USA

Moni Naor
Weizmann Institute of Science, Rehovot, Israel

Oscar Nierstrasz
University of Bern, Switzerland

C. Pandu Rangan
Indian Institute of Technology, Madras, India

Bernhard Steffen
University of Dortmund, Germany

Madhu Sudan
Massachusetts Institute of Technology, MA, USA

Demetri Terzopoulos
University of California, Los Angeles, CA, USA

Doug Tygar
University of California, Berkeley, CA, USA

Gerhard Weikum
Max-Planck Institute of Computer Science, Saarbruecken, Germany
Privacy in statistical databases is a discipline whose purpose is to provide solutions to the tension between the increasing social, political and economical demand of accurate information, and the legal and ethical obligation to protect the privacy of the various parties involved. Those parties are the respondents (the individuals and enterprises to which the database records refer), the data owners (those organizations spending money in data collection) and the users (the ones querying the database, who would like their queries to stay confidential). Beyond law and ethics, there are also practical reasons for data collecting agencies to invest in respondent privacy: if individual respondents feel their privacy guaranteed, they are likely to provide more accurate responses. Data owner privacy is primarily motivated by practical considerations: if an enterprise collects data at its own expense, it may wish to minimize leakage of those data to other enterprises (even to those with whom joint data exploitation is planned). Finally, user privacy results in increased user satisfaction, even if it may curtail the ability of the database owner to profile users.

There are at least two traditions in statistical database privacy, both of which started in the 1970s: one stems from official statistics, where the discipline is also known as statistical disclosure control (SDC), and the other originates from computer science and database technology. In official statistics, the basic concern is respondent privacy. In computer science, one started with respondent privacy but, from 2000 onwards, growing attention has been devoted to owner privacy (privacy-preserving data mining) and user privacy (private information retrieval). In the last few years, the interest and the achievements of computer scientists in the topic have substantially increased.

Privacy in Statistical Databases 2008 (PSD 2008) was held under the sponsorship of the UNESCO Chair in Data Privacy, which intends to act as a stable umbrella for the PSD biennial conference series from now on. PSD 2008 was a successor of PSD 2006, the final conference of the Eurostat-funded CENEX-SDC project, held in Rome in 2006, and PSD 2004, the final conference of the EU FP5 CASC project (IST-2000-25069), held in Barcelona in 2004. Proceedings of PSD 2006 and PSD 2004 were published by Springer in LNCS 4302 and LNCS 3050, respectively. The three PSD conferences held so far are a follow-up of a series of high-quality technical conferences on SDC which started one decade ago with “Statistical Data Protection-SDP 1998”, held in Lisbon in 1998 and with proceedings published by OPOCE, and continued with the AMRADS project SDC Workshop, held in Luxemburg in 2001 and with proceedings published by Springer in LNCS 2316.

For PSD 2008, the Program Committee accepted 27 papers out of 37 submissions from 16 different countries in five different continents. Each submitted paper received at least two reviews. These proceedings contain the revised versions of
the accepted papers, which are a fine blend of contributions from the areas of official statistics and computer science. Covered topics include tabular data protection, methods and case studies for microdata protection, disclosure risk assessment in microdata protection, on-line databases and remote access, privacy-preserving data mining, private information retrieval and legal issues.

We are indebted to many people. First, to the Government of Catalonia for financial support of the UNESCO Chair in Data Privacy, which enabled the latter to sponsor PSD 2008. Also, to the Organizing Committee for making the conference possible and especially to Jesús Manjón, who helped with these proceedings. In evaluating the papers we were assisted by the Program Committee and the following external reviewers: Javier Herranz, Marlow Lemons, Thomas B. Pedersen, Adam Persing and Elizabeth Ransom.

We also wish to thank all the authors of submitted papers and apologize for possible omissions.

July 2008

Josep Domingo-Ferrer
Yücel Saygın
Organization

Program Committee

John Abowd  Cornell University, USA
Elisa Bertino  CERIAS, Purdue University, USA
Jordi Castro  Polytechnical University of Catalonia, Spain
Lawrence Cox  Nat. Center for Health Statistics, USA
Josep Domingo-Ferrer  Rovira i Virgili University, Catalonia, Spain
Mark Elliot  Manchester University, UK
Elena Ferrari  University of Insubria, Italy
Stephen Fienberg  Carnegie Mellon University, USA
Luisa Franconi  ISTAT, Italy
Sarah Giessing  Destatis, Germany
Anco Hundepool  Statistics Netherlands, The Netherlands
Ramayya Krishnan  Carnegie Mellon University, USA
Julia Lane  NORC/University of Chicago, USA
Jane Longhurst  Office for National Statistics, UK
Bradley Malin  Vanderbilt University, USA
Josep M. Mateo-Sanz  Rovira i Virgili University, Catalonia, Spain
Krish Muralidhar  University of Kentucky, USA
Jean-Marc Museux  EUROSTAT, European Union
Silvia Polettini  University of Naples, Italy
Yosef Rinott  Hebrew University, Israel
Gerd Ronning  University of Tübingen, Germany
Juan José Salazar  University of La Laguna, Spain
Yücel Saygın  Sabancı University, Turkey
Eric Schulte-Nordholt  Statistics Netherlands, The Netherlands
Francesc Sebé  Rovira i Virgili University, Catalonia, Spain
Natalie Shlomo  University of Southampton, UK
Julian Stander  University of Plymouth, UK
Vicenç Torra  IIIA-CSIC, Catalonia, Spain
William E. Winkler  Census Bureau, USA
Laura Zayatz  Census Bureau, USA

Program Chair

Josep Domingo-Ferrer  Rovira i Virgili University, Catalonia, Spain

General Chair

Yücel Saygın  Sabancı University, Turkey
VIII Organization

Organizing Committee

Jordi Castellà-Roca  Rovira i Virgili University, Catalonia, Spain
Aysegul Cayci  Sabancı University, Turkey
Ercument Cicek  Sabancı University, Turkey
Aras Sami Kubilay  Sabancı University, Turkey
Jesús Manjón  Rovira i Virgili University, Catalonia, Spain
Antoni Martínez-Ballester  Rovira i Virgili University, Catalonia, Spain
Glòria Pujol  Rovira i Virgili University, Catalonia, Spain
# Table of Contents

## Tabular Data Protection

Using a Mathematical Programming Modeling Language for Optimal CTA ........................................................... 1  
*Jordi Castro and Daniel Baena*

A Data Quality and Data Confidentiality Assessment of Complementary Cell Suppression ........................................ 13  
*Lawrence H. Cox*

Pre-processing Optimisation Applied to the Classical Integer Programming Model for Statistical Disclosure Control .............. 24  
*Martin Serpell, Alistair Clark, Jim Smith, and Andrea Staggemeier*

How to Make the $\tau$-ARGUS Modular Method Applicable to Linked Tables ............................................................... 37  
*Peter-Paul de Wolf and Sarah Giessing*

Bayesian Assessment of Rounding-Based Disclosure Control ............ 50  
*Jon J. Forster and Roger C. Gill*

Cell Bounds in Two-Way Contingency Tables Based on Conditional Frequencies .......................................................... 64  
*Byran Smucker and Aleksandra B. Slavković*

Invariant Post-tabular Protection of Census Frequency Counts .......... 77  
*Natalie Shlomo and Caroline Young*

## Microdata Protection: Methods and Case Studies

A Practical Approach to Balancing Data Confidentiality and Research Needs: The NHIS Linked Mortality Files .......................... 90  
*Kimberly Lochner, Stephanie Bartee, Gloria Wheatcroft, and Christine Cox*

From $t$-Closeness to PRAM and Noise Addition Via Information Theory ................................................................. 100  
*David Rebollo-Monedero, Jordi Forné, and Josep Domingo-Ferrer*

Robustification of Microdata Masking Methods and the Comparison with Existing Methods ........................................... 113  
*Matthias Templ and Bernhard Meindl*
A Preliminary Investigation of the Impact of Gaussian Versus t-Copula for Data Perturbation ............................................ 127
  Mario Trottini, Krish Muralidhar, and Rathindra Sarathy

Anonymisation of Panel Enterprise Microdata – Survey of a German Project .......................................................... 139
  Maurice Brandt, Rainer Lenz, and Martin Rosemann

Microdata Protection: Disclosure Risk Assessment

Towards a More Realistic Disclosure Risk Assessment 152
  Jordi Nin, Javier Herranz, and Vicenç Torra

Assessing Disclosure Risk for Record Linkage 166
  Chris Skinner

Robust Statistics Meets SDC: New Disclosure Risk Measures for Continuous Microdata Masking 177
  Matthias Templ and Bernhard Meindl

Parallelizing Record Linkage for Disclosure Risk Assessment 190
  Joan Guisado-Gámez, Arnaud Prat-Pérez, Jordi Nin, Victor Muntés-Mulero, and Josep Ll. Larriba-Pey

Extensions of the Re-identification Risk Measures Based on Log-Linear Models 203
  Daniela Ichim

Use of Auxiliary Information in Risk Estimation 213
  Loredana Di Consiglio and Silvia Polettini

Accounting for Intruder Uncertainty Due to Sampling When Estimating Identification Disclosure Risks in Partially Synthetic Data 227
  Jörg Drechsler and Jerome P. Reiter

How Protective Are Synthetic Data? 239
  John M. Abowd and Lars Vilhuber

On-Line Databases and Remote Access

Auditing Categorical SUM, MAX and MIN Queries 247
  Francesco M. Malvestuto

Reasoning under Uncertainty in On-Line Auditing 257
  Gerardo Canfora and Bice Cavallo

A Remote Analysis Server – What Does Regression Output Look Like? 270
  Christine M. O’Keefe and Norm M. Good
<table>
<thead>
<tr>
<th>Title</th>
<th>Page</th>
</tr>
</thead>
<tbody>
<tr>
<td>Privacy-Preserving Data Mining and Private Information Retrieval</td>
<td></td>
</tr>
<tr>
<td>Accuracy in Privacy-Preserving Data Mining Using the Paradigm of</td>
<td>284</td>
</tr>
<tr>
<td>Cryptographic Elections</td>
<td></td>
</tr>
<tr>
<td>Emmanouil Magkos, Manolis Maragoudakis, Vassilis Chrissikopoulos,</td>
<td></td>
</tr>
<tr>
<td>and Stefanos Gridzalis</td>
<td></td>
</tr>
<tr>
<td>A Privacy-Preserving Framework for Integrating Person-Specific</td>
<td>298</td>
</tr>
<tr>
<td>Databases</td>
<td></td>
</tr>
<tr>
<td>Murat Kantarcioglu, Wei Jiang, and Bradley Malin</td>
<td></td>
</tr>
<tr>
<td>Peer-to-Peer Private Information Retrieval</td>
<td>315</td>
</tr>
<tr>
<td>Josep Domingo-Ferrer and Maria Bras-Amorós</td>
<td></td>
</tr>
<tr>
<td>Legal Issues</td>
<td></td>
</tr>
<tr>
<td>Legal, Political and Methodological Issues in Confidentiality in the</td>
<td></td>
</tr>
<tr>
<td>European Statistical System</td>
<td>324</td>
</tr>
<tr>
<td>Jean Marc Museux, Martine Peeters, and Maria João Santos</td>
<td></td>
</tr>
<tr>
<td>Author Index</td>
<td>335</td>
</tr>
</tbody>
</table>