

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

TU Dortmund University, Germany

Madhu Sudan

Microsoft Research, Cambridge, MA, USA

Demetri Terzopoulos

University of California, Los Angeles, CA, USA

Doug Tygar

University of California, Berkeley, CA, USA

Gerhard Weikum

Max Planck Institute for Informatics, Saarbruecken, Germany

Josep Domingo-Ferrer
Emmanouil Magkos (Eds.)

Privacy in Statistical Databases

UNESCO Chair in Data Privacy
International Conference, PSD 2010
Corfu, Greece, September 22-24, 2010
Proceedings

Volume Editors

Josep Domingo-Ferrer
Universitat Rovira i Virgili
Department of Computer Engineering and Mathematics
UNESCO Chair in Data Privacy
Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain
E-mail: josep.domingo@urv.cat

Emmanouil Magkos
Ionian University, Department of Informatics
Plateia Tsirigoti 7, 49100 Corfu, Greece
E-mail: emagos@ionio.gr

Library of Congress Control Number: 2010934215

CR Subject Classification (1998): H.2, D.4.6, K.6.5, C.2, E.3, E.1

LNCS Sublibrary: SL 3 – Information Systems and Application, incl. Internet/Web and HCI

ISSN	0302-9743
ISBN-10	3-642-15837-4 Springer Berlin Heidelberg New York
ISBN-13	978-3-642-15837-7 Springer Berlin Heidelberg New York

This work is subject to copyright. All rights are reserved, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, re-use of illustrations, recitation, broadcasting, reproduction on microfilms or in any other way, and storage in data banks. Duplication of this publication or parts thereof is permitted only under the provisions of the German Copyright Law of September 9, 1965, in its current version, and permission for use must always be obtained from Springer. Violations are liable to prosecution under the German Copyright Law.

springer.com

© Springer-Verlag Berlin Heidelberg 2010
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper 06/3180

Preface

Privacy in statistical databases is a discipline whose purpose is to provide solutions to the tension between the social, political, economic and corporate demand for 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 or the search engine, who would like their queries to stay confidential). Beyond law and ethics, there are also practical reasons for data-collecting agencies and corporations 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: the first one stems from official statistics, where the discipline is also known as statistical disclosure control (SDC), and the second one originates from computer science and database technology. In official statistics, the basic concern is respondent privacy. In computer science, the initial motivation was also 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, as reflected in the contents of this volume.

“Privacy in Statistical Databases 2010” (PSD 2010) was held under the sponsorship of the UNESCO Chair in Data Privacy, which has provided a stable umbrella for the PSD biennial conference series since PSD 2008, held in Istanbul. Previous PSD conferences were 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 2008, PSD 2006 and PSD 2004 were published by Springer in LNCS 5262, LNCS 4302 and LNCS 3050, respectively. The four PSD conferences held so far are a follow-up of a series of high-quality technical conferences on SDC, which started twelve years ago with “Statistical Data Protection-SDP’98”, 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.

The PSD 2010 Program Committee accepted for publication in this volume 25 papers out of 37 full submissions. Furthermore, 5 additional submissions were received which were reviewed for short presentation at the conference and inclusion in the companion CD proceedings. Papers came from 16 different countries from 4 different continents. Each submitted paper received at least two reviews. The revised versions of the 25 accepted papers in this volume are a fine blend of contributions from official statistics and computer science. Topics covered include tabular data protection, microdata protection, synthetic data, differential privacy, on-line databases and remote access, privacy-preserving protocols, and legal issues.

We are indebted to many people. First, to the Government of Catalonia for financial support to the UNESCO Chair in Data Privacy, which enabled the latter to sponsor PSD 2010. Also, to the Organization Committee for making the conference possible and especially to Jesús Manjón, who helped prepare these proceedings. In evaluating the papers we were assisted by the Program Committee and the following external reviewers: Anne-Sophie Charest, Jörg Drechsler, Sara Foresti, Flavio Foschi, Gabriel Ghinita and Peter-Paul de Wolf.

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

July 2010

Josep Domingo-Ferrer
Emmanouil Magkos

Organization

Program Committee

John Abowd	Cornell University, USA
Bettina Berendt	Katholieke Universiteit Leuven, Belgium
Elisa Bertino	CERIAS, Purdue University, USA
Jordi Castro	Polytechnical University of Catalonia, Spain
Lawrence Cox	Nat. Center for Health Statistics, USA
Vassilios Chrissikopoulos	Ionian University, Greece
Josep Domingo-Ferrer	Universitat Rovira i Virgili, Catalonia, Spain
Mark Elliot	Manchester University, UK
Stephen Fienberg	Carnegie Mellon University, USA
Luisa Franconi	ISTAT, Italy
Sarah Gießing	Destatis, Germany
Anco Hundepool	Statistics Netherlands, The Netherlands
Julia Lane	National Science Foundation, USA
Emmanouil Magkos	Ionian University, Greece
Bradley Malin	Vanderbilt University, USA
Josep M. Mateo-Sanz	Universitat Rovira i Virgili, 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
Pierangela Samarati	University of Milan, Italy
Yücel Saygın	Sabancı University, Turkey
Eric Schulte-Nordholt	Statistics Netherlands, The Netherlands
Natalie Shlomo	University of Southampton, UK
Vicenç Torra	IIIA-CSIC, Catalonia, Spain
Vassilios Verykios	University of Thessaly, Greece
William E. Winkler	Census Bureau, USA
Laura Zayatz	Census Bureau, USA

Program Chair

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

General Co-chairs

Emmanouil Magkos	Ionian University, Greece
Vassilios Chrissikopoulos	Ionian University, Greece

Organization Committee

Ioannis Karydis	Ionian University, Greece
Jesús Manjón	Universitat Rovira i Virgili, Catalonia, Spain
Alexandros Panaretos	Ionian University, Greece
Glòria Pujol	Universitat Rovira i Virgili, Catalonia, Spain
Spyros Sioutas	Ionian University, Greece

Table of Contents

Tabular Data Protection

Privacy Disclosure Analysis and Control for 2D Contingency Tables Containing Inaccurate Data	1
<i>Bing Liang, Kevin Chiew, Yingjiu Li, and Yanjiang Yang</i>	
A Tool for Analyzing and Fixing Infeasible RCTA Instances	17
<i>Jordi Castro and José A. González</i>	
Branch-and-Cut versus Cut-and-Branch Algorithms for Cell Suppression	29
<i>Juan-José Salazar-González</i>	
Data Swapping for Protecting Census Tables	41
<i>Natalie Shlomo, Caroline Tudor, and Paul Groom</i>	
Eliminating Small Cells from Census Counts Tables: Some Considerations on Transition Probabilities	52
<i>Sarah Giessing and Jörg Höhne</i>	
Three Ways to Deal with a Set of Linked SBS Tables Using τ -ARGUS	66
<i>Peter-Paul de Wolf and Anco Hundepool</i>	

Microdata Protection

IPUMS-International Statistical Disclosure Controls: 159 Census Microdata Samples in Dissemination, 100+ in Preparation	74
<i>Robert McCaa, Steven Ruggles, and Matt Sobek</i>	
Uncertainty for Anonymity and 2-Dimensional Range Query Distortion	85
<i>Spyros Sioutas, Emmanouil Magkos, Ioannis Karydis, and Vassilios S. Verykios</i>	
PRAM Optimization Using an Evolutionary Algorithm	97
<i>Jordi Marés and Vicenç Torra</i>	
Multiplicative Noise Protocols	107
<i>Anna Oganian</i>	
Measurement Error and Statistical Disclosure Control	118
<i>Natalie Shlomo</i>	

Semantic Microaggregation for the Anonymization of Query Logs	127
<i>Arnau Erola, Jordi Castellà-Roca, Guillermo Navarro-Arribas, and Vicenç Torra</i>	

Data Environment Analysis and the Key Variable Mapping System	138
<i>Mark Elliot, Susan Lomax, Elaine Mackey, and Kingsley Purdam</i>	

Synthetic Data

Using Support Vector Machines for Generating Synthetic Datasets	148
<i>Jörg Drechsler</i>	

Synthetic Data for Small Area Estimation	162
<i>Joseph W. Sakshaug and Trivellore E. Raghunathan</i>	

Disclosure Risk of Synthetic Population Data with Application in the Case of EU-SILC	174
<i>Matthias Templ and Andreas Alfons</i>	

Differential Privacy

Differential Privacy and the Risk-Utility Tradeoff for Multi-dimensional Contingency Tables	187
<i>Stephen E. Fienberg, Alessandro Rinaldo, and Xiaolin Yang</i>	

Does Differential Privacy Protect Terry Gross' Privacy?	200
<i>Krish Muralidhar and Rathindra Sarathy</i>	

Some Additional Insights on Applying Differential Privacy for Numeric Data	210
<i>Rathindra Sarathy and Krish Muralidhar</i>	

On-Line Databases and Remote Access

Remote Data Access and the Risk of Disclosure from Linear Regression: An Empirical Study	220
<i>Philipp Bleninger, Jörg Drechsler, and Gerd Ronning</i>	

The Microdata Analysis System at the U.S. Census Bureau	234
<i>Jason Lucero and Laura Zayatz</i>	

Establishing an Infrastructure for Remote Access to Microdata at Eurostat	249
<i>Wolf Heinrich Reuter and Jean-Marc Museux</i>	

Privacy-Preserving Protocols

Coprivacy: Towards a Theory of Sustainable Privacy	258
<i>Josep Domingo-Ferrer</i>	
Privacy-Preserving Record Linkage	269
<i>Rob Hall and Stephen E. Fienberg</i>	

Legal Issues

Strategies to Achieve SDC Harmonisation at European Level: Multiple Countries, Multiple Files, Multiple Surveys	284
<i>Daniela Ichim and Luisa Franconi</i>	
Author Index	297