

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

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

Moshe Y. Vardi

Rice University, Houston, TX, USA

Gerhard Weikum

Max-Planck Institute of Computer Science, Saarbruecken, Germany

Josep Domingo-Ferrer Luisa Franconi (Eds.)

Privacy in Statistical Databases

CENEX-SDC Project International Conference, PSD 2006
Rome, Italy, December 13-15, 2006
Proceedings

Volume Editors

Josep Domingo-Ferrer
Rovira i Virgili University of Tarragona
Dept. of Computer Engineering and Mathematics
Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain
E-mail: josep.domingo@urv.cat

Luisa Franconi
ISTAT, Servizio Progettazione e Supporto Metodologico
nei Processi di Produzione Statistica
Via Cesare Balbo 16, 00184 Roma, Italy
E-mail: franconi@istat.it

Library of Congress Control Number: 2006936080

CR Subject Classification (1998): H.2.8, H.2, G.3, K.4.1, I.2.4

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

ISSN	0302-9743
ISBN-10	3-540-49330-1 Springer Berlin Heidelberg New York
ISBN-13	978-3-540-49330-3 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 is a part of Springer Science+Business Media
springer.com

© Springer-Verlag Berlin Heidelberg 2006
Printed in Germany

Typesetting: Camera-ready by author, data conversion by Scientific Publishing Services, Chennai, India
Printed on acid-free paper SPIN: 11930242 06/3142 5 4 3 2 1 0

Preface

Privacy in statistical databases is a discipline whose purpose is to provide solutions to the conflict between the increasing social, political and economical demand of accurate information, and the legal and ethical obligation to protect the privacy of the individuals and enterprises to which statistical data refer. Beyond law and ethics, there are also practical reasons for statistical agencies and data collectors to invest in this topic: if individual and corporate respondents feel their privacy guaranteed, they are likely to provide more accurate responses.

There are at least two traditions in statistical database privacy: 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. Both started in the 1970s, but the 1980s and the early 1990s saw little privacy activity on the computer science side. The Internet era has strengthened the interest of both statisticians and computer scientists in this area. Along with the traditional topics of tabular and microdata protection, some research lines have revived and/or appeared, such as privacy in queryable databases and protocols for private data computation.

Privacy in Statistical Databases 2006 (PSD 2006) was the main conference of the CENEX-SDC project (Center of Excellence in SDC), funded by EUROSTAT (European Commission) and held in Rome, December 13–15, 2006. PSD 2006 is a successor of PSD 2004, the final conference of the CASC project (IST-2000-25069), held in Barcelona in 2004 and with proceedings published by Springer as LNCS vol. 3050. Those two PSD conferences follow a tradition of high-quality technical conferences on SDC which started 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 in Springer LNCS vol. 2316.

The Program Committee accepted 31 papers out of 45 submissions from 17 different countries in four 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 official statistics and computer science. Covered topics include methods for tabular data protection, methods for individual data (microdata) protection, assessments of analytical utility and disclosure risk, protocols for private computation, case studies and SDC software.

We are indebted to many people. First, to EUROSTAT for sponsoring the CENEX project and PSD 2006. Also, to those who made the conference and these proceedings possible: the Organization Committee (Xenia Caruso, Jordi Castellà-Roca, Maurizio Lucarelli, Jesús Manjón, Antoni Martínez-Ballesté and Micaela Paciello). In evaluating the papers we received the help of the Program

Committee and the following external reviewers: Lisa Dragoset, José Antonio González, Krish Muralidhar, Bryan Richetti and Monica Scannapieco.

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

September 2006

Josep Domingo-Ferrer
Luisa Franconi

Privacy in Statistical Databases - PSD 2006

Program Committee

John Abowd (Cornell University and Census Bureau, USA)
Jordi Castro (Polytechnical University of Catalonia)
Lawrence Cox (National Center for Health Statistics, USA)
Ramesh Dandekar (Energy Information Administration, USA)
Josep Domingo-Ferrer (Rovira i Virgili University, Catalonia)
Mark Elliot (Manchester University, UK)
Luisa Franconi (ISTAT, Italy)
Sarah Giessing (Destatis, Germany)
Jobst Heitzig (Destatis, Germany)
Anco Hundepool (Statistics Netherlands)
Ramayya Krishnan (Carnegie Mellon University, USA)
Julia Lane (NORC/University of Chicago, USA)
Jane Longhurst (Office for National Statistics, UK)
Silvia Polettini (University of Naples, Italy)
Gerd Ronning (University of Tübingen, Germany)
Juan José Salazar (University of La Laguna, Spain)
Maria João Santos (EUROSTAT, European Commission)
Eric Schulte-Nordholt (Statistics Netherlands)
Francesc Sebé (Rovira i Virgili University, Catalonia)
Natalie Shlomo (University of Southampton, UK; Hebrew University, Israel)
Chris Skinner (University of Southampton, UK)
Julian Stander (University of Plymouth, UK)
Vicenç Torra (IIIA-CSIC, Catalonia)
William E. Winkler (Census Bureau, USA)

Program Chair

Josep Domingo-Ferrer (Rovira i Virgili University, Catalonia)

General Chair

Luisa Franconi (ISTAT, Italy)

Organization Committee

Xenia Caruso (ISTAT, Italy)
Jordi Castellà-Roca (Rovira i Virgili University, Catalonia)

VIII Organization

Maurizio Lucarelli (ISTAT, Italy)

Jesús Manjón (Rovira i Virgili University, Catalonia)

Antoni Martínez-Ballesté (Rovira i Virgili University, Catalonia)

Micaela Paciello (ISTAT, Italy)

Table of Contents

Methods for Tabular Protection

A Method for Preserving Statistical Distributions Subject to Controlled Tabular Adjustment	1
<i>Lawrence H. Cox, Jean G. Orelieu, Babubhai V. Shah</i>	
Automatic Structure Detection in Constraints of Tabular Data	12
<i>Jordi Castro, Daniel Baena</i>	
A New Approach to Round Tabular Data	25
<i>Juan José Salazar González</i>	
Harmonizing Table Protection: Results of a Study	35
<i>Sarah Giessing, Stefan Dittrich</i>	

Utility and Risk in Tabular Protection

Effects of Rounding on the Quality and Confidentiality of Statistical Data	48
<i>Lawrence H. Cox, Jay J. Kim</i>	
Disclosure Analysis for Two-Way Contingency Tables	57
<i>Haibing Lu, Yingjiu Li, Xintao Wu</i>	
Statistical Disclosure Control Methods Through a Risk-Utility Framework	68
<i>Natalie Shlomo, Caroline Young</i>	
A Generalized Negative Binomial Smoothing Model for Sample Disclosure Risk Estimation	82
<i>Yosef Rinott, Natalie Shlomo</i>	
Entry Uniqueness in Margined Tables	94
<i>Shmuel Onn</i>	

Methods for Microdata Protection

Combinations of SDC Methods for Microdata Protection	102
<i>Anna Oganian, Alan F. Karr</i>	

A Fixed Structure Learning Automaton Micro-aggregation Technique for Secure Statistical Databases	114
<i>Ebaa Fayyoubi, B. John Oommen</i>	
Optimal Multivariate 2-Microaggregation for Microdata Protection: A 2-Approximation	129
<i>Josep Domingo-Ferrer, Francesc Sebé</i>	
Using the Jackknife Method to Produce Safe Plots of Microdata	139
<i>Jobst Heitzig</i>	
Combining Blanking and Noise Addition as a Data Disclosure Limitation Method	152
<i>Anton Flossmann, Sandra Lechner</i>	
Why Swap When You Can Shuffle? A Comparison of the Proximity Swap and Data Shuffle for Numeric Data	164
<i>Krish Muralidhar, Rathindra Sarathy, Ramesh Dandekar</i>	
Adjusting Survey Weights When Altering Identifying Design Variables Via Synthetic Data	177
<i>Robin Mitra, Jerome P. Reiter</i>	

Utility and Risk in Microdata Protection

Risk, Utility and PRAM	189
<i>Peter-Paul de Wolf</i>	
Distance Based Re-identification for Time Series, Analysis of Distances	205
<i>Jordi Nin, Vicenç Torra</i>	
Beyond k -Anonymity: A Decision Theoretic Framework for Assessing Privacy Risk	217
<i>Guy Lebanon, Monica Scannapieco, Mohamed R. Fouad, Elisa Bertino</i>	
Using Mahalanobis Distance-Based Record Linkage for Disclosure Risk Assessment	233
<i>Vicenç Torra, John M. Abowd, Josep Domingo-Ferrer</i>	
Improving Individual Risk Estimators	243
<i>Loredana Di Consiglio, Silvia Polettini</i>	

Protocols for Private Computation

Single-Database Private Information Retrieval Schemes : Overview, Performance Study, and Usage with Statistical Databases	257
<i>Carlos Aguilar Melchor, Yves Deswarte</i>	
Privacy-Preserving Data Set Union	266
<i>Alberto Maria Segre, Andrew Wildenberg, Veronica Vieland, Ying Zhang</i>	
“Secure” Log-Linear and Logistic Regression Analysis of Distributed Databases	277
<i>Stephen E. Fienberg, William J. Fulp, Aleksandra B. Slavkovic, Tracey A. Wrobel</i>	

Case Studies

Measuring the Impact of Data Protection Techniques on Data Utility: Evidence from the Survey of Consumer Finances	291
<i>Arthur Kennickell, Julia Lane</i>	
Protecting the Confidentiality of Survey Tabular Data by Adding Noise to the Underlying Microdata: Application to the Commodity Flow Survey	304
<i>Paul Massell, Laura Zayatz, Jeremy Funk</i>	
Italian Household Expenditure Survey: A Proposal for Data Dissemination	318
<i>Mario Trottini, Luisa Franconi, Silvia Polettini</i>	

Software

The ARGUS Software in CENEX	334
<i>Anco Hundepool</i>	
Software Development for SDC in R	347
<i>Mario Templ</i>	
On Secure e-Health Systems	360
<i>Milan Marković</i>	
IPUMS-International High Precision Population Census Microdata Samples: Balancing the Privacy-Quality Tradeoff by Means of Restricted Access Extracts	375
<i>Robert McCaa, Steven Ruggles, Michael Davern, Tami Swenson, Krishna Mohan Palipudi</i>	

Author Index	383
---------------------------	------------