Lecture Notes in Computer Science 3050

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

Editorial Board

Springer
Berlin
Heidelberg
New York
Hong Kong
London
Milan
Paris
Tokyo

Josep Domingo-Ferrer Vicenç Torra (Eds.)

Privacy in Statistical Databases

CASC Project Final Conference, PSD 2004
Barcelona, Catalonia, Spain, June 9-11, 2004
Proceedings

Springer

Volume Editors

Josep Domingo-Ferrer
Universitat Rovira i Virgili
Department of Computer Engineering and Mathematics
Av. Països Catalans 26, 43007 Tarragona, Catalonia, Spain
E-mail: jdomingo@etse.urv.es

Vicenç Torra
Institut d'Investigació en Intel·ligència Artificial
Campus de Bellaterra, 08193 Bellaterra, Catalonia, Spain
E-mail: vtorra@iiia.csic.es

Library of Congress Control Number: 2004106913

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

ISSN 0302-9743
ISBN 3-540-22118-2 Springer-Verlag Berlin Heidelberg New York

Springer-Verlag is a part of Springer Science+Business Media

springeronline.com

Printed in Germany

Typesetting: Camera-ready by author, data conversion by Olgun Computergrafik
Printed on acid-free paper SPIN: 11008736 06/3142 5 4 3 2 1 0

Preface

Privacy in statistical databases is about finding tradeoffs to the tension between the increasing societal and economical demand for accurate information and the legal and ethical obligation to protect the privacy of individuals and enterprises, which are the source of the statistical data. Statistical agencies cannot expect to collect accurate information from individual or corporate respondents unless these feel the privacy of their responses is guaranteed; also, recent surveys of Web users show that a majority of these are unwilling to provide data to a Web site unless they know that privacy protection measures are in place.

"Privacy in Statistical Databases 2004" (PSD 2004) was the final conference of the CASC project ("Computational Aspects of Statistical Confidentiality", IST-2000-25069). PSD 2004 is in the style of the following conferences: "Statistical Data Protection", held in Lisbon in 1998 and with proceedings published by the Office of Official Publications of the EC, and also the AMRADS project SDC Workshop, held in Luxemburg in 2001 and with proceedings published by Springer-Verlag, as LNCS Vol. 2316.

The Program Committee accepted 29 papers out of 44 submissions from 15 different countries on four continents. Each submitted paper received at least two reviews. These proceedings contain the revised versions of the accepted papers. These papers cover the foundations and methods of tabular data protection, masking methods for the protection of individual data (microdata), synthetic data generation, disclosure risk analysis, and software/case studies.

Many people deserve our gratitude. The conference and these proceedings would not have existed without the Organization Chair, Enric Ripoll, and the Organizing Committee (Jordi Castellà, Antoni Martínez, Francesc Sebé and Julià Urrutia). In evaluating the papers submitted we received the help of the Program Committee and four external reviewers (Jörg Höhne, Silvia Polettini, Yosef Rinott and Giovanni Seri).

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

March 2004

Josep Domingo-Ferrer
Vicenç Torra

Privacy in Statistical Databases – PSD 2004

Program Committee

Rakesh Agrawal (IBM Almaden, USA)
David Brown (ONS, UK)
Jordi Castro (Polytechnical University of Catalonia)
Lawrence Cox (Nat. Center for Health Statistics, USA)
Ramesh Dandekar (EIA, USA)
Stephen Fienberg (Carnegie Mellon University, USA)
Luisa Franconi (ISTAT, Italy)
Sarah Giessing (Destatis, Germany)
Anco Hundepool (Statistics Netherlands, The Netherlands)
Julia Lane (Urban Institute, USA)
Josep M. Mateo-Sanz (Universitat Rovira i Virgili, Catalonia)
William E. Winkler (Census Bureau, USA)
Laura Zayatz (Census Bureau, USA)

Program Chairs

Josep Domingo-Ferrer (Universitat Rovira i Virgili, Catalonia)
Vicenç Torra (IIIA-CSIC, Catalonia)

Organization Committee

Jordi Castellà Roca (Universitat Rovira i Virgili, Catalonia)
Antoni Martínez-Ballesté (Universitat Rovira i Virgili, Catalonia)
Francesc Sebé (Universitat Rovira i Virgili, Catalonia)
Julià Urrutia (IDESCAT, Catalonia)

Organization Chair

Enric Ripoll (IDESCAT, Catalonia)

Table of Contents

Foundations of Tabular Protection

Methods for Tabular Protection

Masking for Microdata Protection

Risk in Microdata Protection

Synthetic Data

Software and Case Studies