

Studies in Computational Intelligence 567

Guillermo Navarro-Arribas
Vicenç Torra *Editors*

Advanced Research in Data Privacy

 Springer

Volume 567

Studies in Computational Intelligence

Series Editor

Janusz Kacprzyk

About this Series

The series “Studies in Computational Intelligence” (SCI) publishes new developments and advances in the various areas of computational intelligence—quickly and with a high quality. The intent is to cover the theory, applications, and design methods of computational intelligence, as embedded in the fields of engineering, computer science, physics and life sciences, as well as the methodologies behind them. The series contains monographs, lecture notes and edited volumes in computational intelligence spanning the areas of neural networks, connectionist systems, genetic algorithms, evolutionary computation, artificial intelligence, cellular automata, self-organizing systems, soft computing, fuzzy systems, and hybrid intelligent systems. Of particular value to both the contributors and the readership are the short publication timeframe and the world-wide distribution, which enable both wide and rapid dissemination of research output.

More information about this series at <http://www.springer.com/series/7092>

Editors

Guillermo Navarro-Arribas and Vicenç Torra

Advanced Research in Data Privacy

 Springer

Editors

Guillermo Navarro-Arribas

Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, Catalonia, Spain

Vicenç Torra

Institut d'Investigació en Intel·ligència Artificial, Consejo Superior de Investigaciones Científicas Campus de la UAB, Catalonia, Spain

ISSN 1860-949X e-ISSN 1860-9503

ISBN 978-3-319-09884-5 e-ISBN 978-3-319-09885-2

DOI 10.1007/978-3-319-09885-2

Springer Cham Heidelberg New York Dordrecht London

Library of Congress Control Number: 2014947701

© Springer International Publishing Switzerland 2015

This work is subject to copyright. All rights are reserved by the Publisher, whether the whole or part of the material is concerned, specifically the rights of translation, reprinting, reuse of illustrations, recitation, broadcasting, reproduction on microfilms or in any other physical way, and transmission or information storage and retrieval, electronic adaptation, computer software, or by similar or dissimilar methodology now known or hereafter developed. Exempted from this legal reservation are brief excerpts in connection with reviews or scholarly analysis or material supplied specifically for the purpose of being entered and executed on a computer system, for exclusive use by the purchaser of the work. Duplication of this publication or parts thereof is permitted only under the provisions of the Copyright Law of the Publisher's location, in its current version, and permission for use must always be obtained from Springer. Permissions for use may be obtained through RightsLink at the Copyright Clearance Center. Violations are liable to prosecution under the respective Copyright Law.

The use of general descriptive names, registered names, trademarks, service marks, etc. in this publication does not imply, even in the absence of a specific statement, that such names are exempt from the relevant protective laws and regulations and therefore free for general use.

While the advice and information in this book are believed to be true and accurate at the date of publication, neither the authors nor the editors nor the publisher can accept any legal responsibility for any errors or omissions that may be made. The publisher

makes no warranty, express or implied, with respect to the material contained herein.

Printed on acid-free paper

Springer is part of Springer Science+Business Media (www.springer.com)

Preface

This book presents the research work done under the auspices of the ARES Project (CSD2007-00004). ARES, which stands for Advanced Research in Privacy and Security, has been one of the most ambitious research projects on computer security and privacy funded by the Spanish Government. It is part of the now extinct CONSOLIDER INGENIO 2010 program, a highly competitive program that aimed to advance knowledge and open new research lines among top Spanish research groups.

The ARES project, coordinated by Josep Domingo-Ferrer from Universitat Rovira i Virgili, started in 2007 and was composed of six research groups from six different institutions: Universitat Rovira i Virgili, Consejo Superior de Investigaciones Científicas, Universidad de Málaga, Universitat Oberta de Catalunya, Universitat Politècnica de Barcelona, and Universitat de les Illes Balears. After 7 years, the project is about to conclude this September 2014. It has given important and internationally recognized results in the areas of computer security and privacy, has significantly increased the research production, and has fueled technology transfer activities.

Among the ARES project, privacy has played an important role. Our group led the work package about privacy within the project, for which Vicenç Torra was mainly responsible. Our intention with this book is to provide a guide to the research done within the ARES project in relation to privacy.

Participants of the project were invited to submit a chapter on their contribution to data privacy and privacy enhancing technologies. These submissions were handled through a peer-review process ending in the current chapters that form the book. In addition, there are three introductory chapters: one that introduces the book, and two introducing the work of the main groups on privacy within ARES. At least one author of each contribution is, or has been, in the ARES project.

This is not an exhaustive enumeration of the work done in the ARES project related to privacy. Instead of giving an exhaustive list we opted for allowing the contributors to actually choose the work that they feel was more relevant and interesting for future research. This book aims to introduce and spread the work done in ARES directly related to privacy. We think it also serves as a review of the current research trends in privacy and privacy enhancing technologies.

We would like to thank all the authors and reviewers that have kindly contributed to this book, as well as Prof. J. Kacprzyk for his support on publishing this book, and the editorial team at Springer for their help.

Guillermo Navarro-Arribas

Vicenç Torra

Bellaterra

June 2014

Contents

Part I Introduction

Advanced Research on Data Privacy in the ARES Project

Guillermo Navarro-Arribas and Vicenç Torra

Selected Privacy Research Topics in the ARES Project: An Overview

Jesús A. Manjón and Josep Domingo-Ferrer

Data Privacy: A Survey of Results

Vicenç Torra and Guillermo Navarro-Arribas

Part II Respondent Privacy: SDC and PPDM

A Review of Attribute Disclosure Control

Stan Matwin, Jordi Nin, Morvarid Sehatkar and Tomasz Szapiro

Data Privacy with R

Daniel Abril, Guillermo Navarro-Arribas and Vicenç Torra

Optimisation-Based Study of Data Privacy by Using PRAM

Jordi Marés, Vicenç Torra and Natalie Shlomo

Part III Respondent Privacy: Semantic Related Respondent Privacy Protection

Semantic Anonymisation of Categorical Datasets

Sergio Martínez, Aida Valls and David Sánchez

Contributions on Semantic Similarity and Its Applications to Data Privacy

Montserrat Batet and David Sánchez

An Information Retrieval Approach to Document Sanitization

David F. Nettleton and Daniel Abril

Part IV Respondent Privacy: Location Privacy

Privacy for LBSs: On Using a Footprint Model to Face the Enemy

Mauro Conti, Roberto Di Pietro and Luciana Marconi

Privacy in Spatio-Temporal Databases: A Microaggregation-Based Approach

Rolando Trujillo-Rasua and Josep Domingo-Ferrer

A Prototype for Anonymizing Trajectories from a Time Series Perspective

Sergi Martínez-Bea

Part V Respondent Privacy: Social Networks

A Summary of k -Degree Anonymous Methods for Privacy-Preserving on Networks

Jordi Casas-Roma, Jordi Herrera-Joancomartí and Vicenç Torra

Evaluating Privacy Risks in Social Networks from the User's Perspective

Michal Sramka

Part VI Respondent Privacy: Other Respondent Privacy Enhancing Technologies

Trustworthy Video Surveillance: An Approach Based on Guaranteeing Data Privacy

Antoni Martínez-Ballesté, Agusti Solanas and Hatem A. Rashwan

Electronic Ticketing: Requirements and Proposals Related to Transport

M. Magdalena Payeras-Capellà, Macià Mut-Puigserver, Josep-Lluís Ferrer-Gomila, Jordi Castellà-Roca and Arnau Vives-Guasch

Security and Privacy Concerns About the RFID Layer of EPC Gen2 Networks

Joaquín García-Alfaro, Jordi Herrera-Joancomartí and Joan Melià-Seguí

Privacy on Mobile Coupons Booklets

M. Francisca Hinarejos, Andreu Pere Isern-Deyà and Josep-Lluís Ferrer-Gomila

Smart User Authentication for an Improved Data Privacy

Vanesa Daza and Matteo Signorini

Part VII User Privacy: Web Search Engines

Multi-party Methods for Privacy-Preserving Web Search: Survey and Contributions

Cristina Romero-Tris, Alexandre Viejo and Jordi Castellà-Roca

DisPA: An Intelligent Agent for Private Web Search

Marc Juárez and Vicenç Torra

A Survey on the Use of Combinatorial Configurations for Anonymous Database Search

Klara Stokes and Maria Bras-Amorós

Part VIII User Privacy: Recommender and Personalized Systems

Privacy-Enhancing Technologies and Metrics in Personalized Information Systems

Javier Parra-Arnau, David Rebollo-Monedero and Jordi Forné

Managing Privacy in the Internet of Things: DocCloud, a Use Case

Juan Vera del Campo, Josep Pegueroles, Juan Hernández Serrano and Miguel Soriano



Part I

Introduction

Advanced Research on Data Privacy in the ARES Project

Guillermo Navarro-Arribas¹✉ and Vicenç Torra²✉

- (1) Department of Information and Communications Engineering, Universitat Autònoma de Barcelona, 08193, Bellaterra, Catalonia, Spain
- (2) Institut d'Investigació en Intel·ligència Artificial, Consejo Superior de Investigaciones Científicas Campus de la UAB, 08193, Bellaterra, Catalonia, Spain

✉ **Guillermo Navarro-Arribas (Corresponding author)**

Email: gnavarro@deic.uab.cat

✉ **Vicenç Torra**

Email: vtorra@iia.csic.es

Email: vtorra@ieee.org

Abstract

Privacy has become an important concern in today's society. The advancement and pervasiveness of information and communication technologies have a great positive impact in our society, they greatly affect how we socialize, the way we do business, or even our individual and social freedom.

1 Introduction

Privacy has become an important concern in today's society. The advancement and pervasiveness of information and communication technologies have a great positive impact in our society, they greatly affect how we socialize, the way we do business, or even our individual and social freedom. At the same time, these new technologies are enabling an unparalleled invasion of privacy. There has been a relatively recent awareness regarding government mass surveillance programs [13], important information leakages in corporation environments [6], or even highly publicized scandals arousing when poorly anonymized user data is made public [4, 25]. Governments, users, and corporations are starting to take privacy seriously. As an example of user awareness, a recent survey from Mozilla identifies privacy as the top priority of users from all regions concerning the future of the Web [24]. All this

interest has motivated an increased interest from the research community in data privacy and privacy enhancing technologies. From more traditional disclosure control techniques rooted in the statistics community to more recent studies in social networks or data mining. The increase in research work from private and public sectors has driven the raise of funded research projects and academic production.

A relevant academic project funded by the Spanish Government is about to conclude in September 2014. The project was called ARES, Advanced Research on Information Security and Privacy, and, as expected, an important part of the project was to advance the research on data privacy.

In this book we give an overall picture of the most notable research work that has come out from the ARES project in relation to data privacy. All the works presented here, have been done under the project umbrella. At the same time these works do provide an state of the art picture of data privacy research, since they include important contributions already recognized in prestigious international conferences and journals.

In the following section we describe the book contents to give an idea of what the reader will find in the rest of book.

Table 1 Groups and their principal investigators taking part in the ARES project

Group	P.I.	Web page
CRISES/URV	Josep Domingo-Ferrer	http://crises2-deim.urv.cat/
IF-PAD/CSIC	Vicenç Torra	http://www.iiia.csic.es/~vtorra/ares/
KISON/UOC	Jordi Herrera-Joancomartí, David Megias	http://kison.uoc.edu
ISG/UPC	Miguel Soriano	http://isg.upc.edu/
GIDET/UIB	Josep Lluís Ferrer-Gomila	http://secom.uib.es/
GSI/UMA	Javier Lopez	https://www.nics.uma.es/

2 The ARES Project and Data Privacy

The ARES project was part of the, currently extinguished, CONSOLIDER INGENIO 2010 program, possibly the most ambitious and competitive research program in Spain. The project is composed of six Spanish research groups in the area of information security and privacy. It started in October 2007 and will end in September 2014.

The specific groups that took part in the project are:

- CRISES/URV: Secure Electronic Commerce group at Universitat Rovira i Virgili of Tarragona.
- IF-PAD/CSIC: IF-PAD, Information Fusion for Privacy and Decision group, located at the Artificial Intelligence Research Institute of CSIC.
- KISON/UOC: K-ryptography and Information Security for Open Networks group from the Universitat Oberta de Catalunya.
- ISG/UPC: Information Security Group at the Universitat Politècnica de Catalunya, in Barcelona.

- GIDET/UIB: the Interdisciplinary Group on Law and Telematics at the Universitat de les Illes Balears.
- GSI/UMA: the Network Information and Computer Security Lab (NICS), former Information Security Group (GSI), from the Universidad de Málaga.

A summary of the groups, the principal investigator, and web page for each group is given in Table 1. The project coordinator has been Josep Domingo-Ferrer from the CRISES/URV group.

The aim of the project was to create technologies to conciliate security, privacy and functionality in the information society.

More precisely, the research work of the project has been settle around three intertwined applications scenarios plus two transversal underpinning areas:

- Scenario 1: protection of critical information infrastructures.
- Scenario 2: ubiquitous computing.
- Scenario 3: secure electronic commerce and digital content distribution.
- Underpinning area 1: data privacy technologies.
- Underpinning area 2: technical-legal issues.

Each gives up to a specific workpackage, two of them transversal: data privacy technologies and technical-legal issues. Two more workpackages are intended for management and for field trial and dissemination. Table 2 lists the workpackages with the group that led each one, and Fig. 1 summarizes the workpackages and their interrelation. A links from WP_i to WP_j means that work done in WP_i is used by WP_j .

Table 2 Workpackages of the ARES project with its leader group

WP1	Critical infrastructure protection	GSI/UMA
WP2	Ubiquitous computing	ISG/UPC
WP3	Secure e-commerce and digital content distribution	KISON/UOC
WP4	Data privacy technologies	IF-PAD/CSIC
WP5	Technical-legal issues	GIDET/UIB
WP6	Field trial, technology transfer, and dissemination	CRISES/URV
WP7	Management	CRISES/URV

As shown, research on privacy is gathered in an specific work package within the project. The WP4, Data privacy technologies, is a transversal worpackage which has been leaded by the IF-PAD/CSIC group. Its main (broad) objective has been to develop or adapt privacy technologies to solve privacy problems aroused form other parts of the projects.

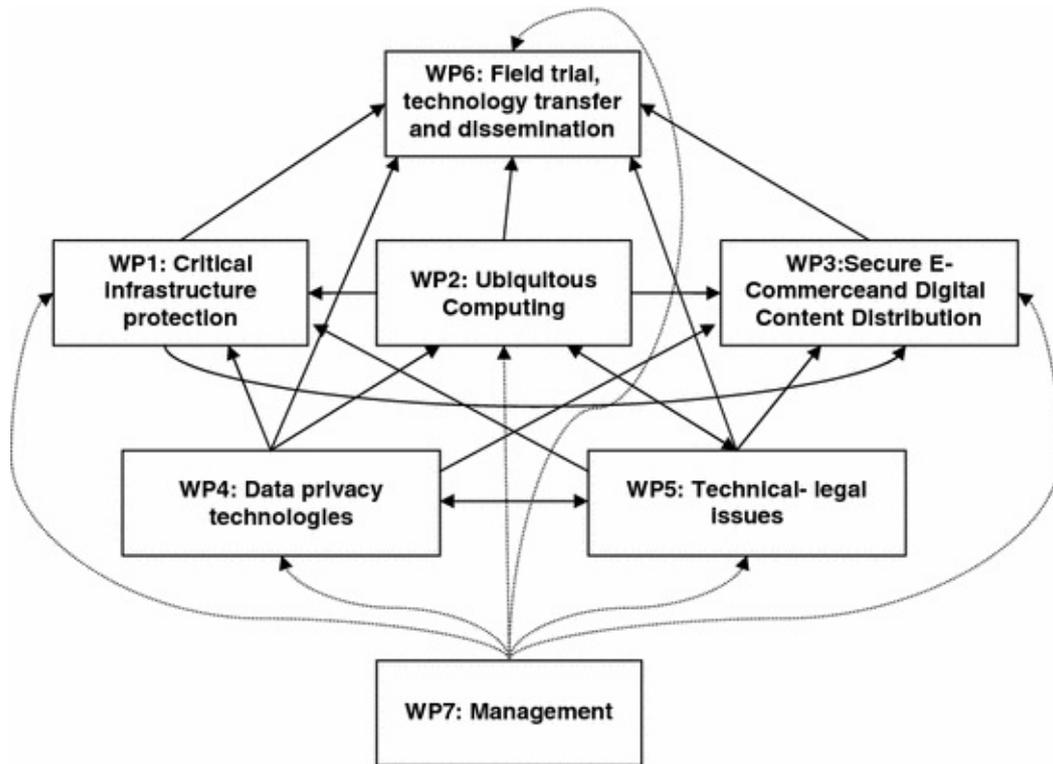


Fig. 1 ARES workpackages

As we will see, the work presented through the book develops several topics and provides an overview of the research carried by the project participants. Some of the presented works include collaborations with researchers outside the project.

3 Data Privacy

Data privacy and privacy enhancing technologies (PET) provide a broad research line which spreads through several specific fields. The common aim is to study the protection of information in order to avoid unintentional disclosure of sensitive information. One of the first disciplines to deal with this problem is Statistical Disclosure Control (SDC). SDC, rooted in the statistics community, develops methods to allow the publication of data from statistical agencies while preserving the privacy of their users. In the case of computer science, data privacy has become an important research line. Under privacy enhancing technologies we find from cryptography to privacy-preserving data mining or private information protocols, ...Although this comprises several disciplines all share the same goal and even some of the techniques, methods and definitions.

The classification of privacy enhancing technologies is difficult due to the overlap in most of the disciplines. Nevertheless a well accepted dimension to classify them is determined by considering whose privacy is being sought [10]. That is, techniques can be classified in terms of whose privacy are attempting to protect. We have thus:

- **Respondent privacy** The respondent is the subject to whom the data is referring to. In the context of SDC, for example in a Census database, the respondents are the concrete subjects included in the Census. The respondent is considered a

passive subject, who cannot act within the system to protect its own data.

- **Owner privacy** The owner is the proprietary or administrator holder of the data. It is normally the one liable for disclosure of sensitive information.
- **User privacy** The user can be seen as the active counterpart of a respondent. That is, the user is a subject that can actively participate in the protection of its own private data. This is usually done through the interaction of the subject with the system.

We have used the term subject to denote any entity taking part in the system. Although subjects will usually be individuals, other types of entities can be accommodated in the previous dimension. Examples of subjects can be: human beings, organizations, computer processes, electronic devices, ...

As we will see all the works presented in this book are either respondent or user privacy approaches. Owner privacy, although important, is rarely found in common scenarios.

Besides respondent, owner, and user, other classification can be found in the literature [2, 11, 12, 16, 26, 35, 37, 40, 41]. For instance it is common to differentiate data privacy methods by their intended use, or by the source of the data to be protected. We have however opted for the already mentioned classification. As we will see, in some cases the classification of an specific work into a concrete class is somewhat fuzzy, since some methods and technologies can address the protection of more than one type of entity.

In the following sections we describe the chapters of the book organized according to the previous classification.

Together with this introduction, there are two introductory chapters that summarize the work carried out by two of the groups of the project. These are the groups with a stronger presence in privacy technologies within the ARES project.

Manjón and Domingo-Ferrer [18] (Chap. 2) describe the work carried out by the group from the Universitat Rovira i Virgili within the ARES project. This group and his leader, Josep Domingo-Ferrer, have been the main coordinator of the ARES project, with a strong participation in the whole project. This chapter allows the reader to get a picture of the work performed by the URV group regarding privacy technologies. Note that some of the works described involve the participation of other groups from the project, and external collaborators.

Torra and Navarro-Arribas [36] (Chap. 3) describe the work done by the IIIA-CSIC group regarding data privacy within the ARES project. This group and his leader Vicenç Torra, have been in charge of leading the workpackage WP4 on data privacy technologies (see Sect. 2).

4 Respondent Privacy

Respondent privacy is possibly the most common case found in data privacy scenarios. Most common Statistical Disclosure Control (SDC) and Privacy-preserving Data Mining (PPDM) techniques usually fit in this category. We present in this book several state of the art works concerned with respondent privacy in several application scenarios.

4.1 SDC and PPDM

Statistical Disclosure Control and Privacy-preserving Data Mining are possibly the most classical works on data privacy. Their common objective is to protect a dataset so statistical analysis and data mining techniques can be applied while preserving the privacy of the respondents.

Matwin et al. [23] (Chap. 4) provide an introductory review of data privacy from statistical disclosure control (SDC) and privacy-preserving data mining (PPDM). Authors make specific emphasis in attribute disclosure control.

Abril et al. [1] (Chap. 5) provide an introduction to privacy-preserving data mining (PPDM) with R. R has become an important language and environment for statistics and data mining and thus it is very well suited for SDC and PPDM. This chapter serves as an introduction to PPDM protection techniques, and information loss and disclosure risk, outlining tools and procedures in R to help introducing practitioners to this field.

Marés et al. [19] (Chap. 6) study the problem of finding optimal transition matrices for Post Randomization Methods (PRAM). PRAM is a method commonly used in SDC to introduce perturbation by using a Markov probability transition matrix. The authors introduce a method based in genetic algorithms to find the optimal matrix. That is, the one with a better balance between disclosure risk and utility.

4.2 Semantic Related Respondent Privacy Protection

The following chapters, depart from the traditional approach of SDC and PPDM to deal with textual data. Traditional SDC and PPDM methods usually deal with numeric or categorical data. In the later case they rely at most in a predefined generalization tree. Recent work has been made to deal with textual data by considering its semantics in a broad sense. This allows to deal with free text, or categorical data without predefined categories. These three chapters show the use of semantic based protection techniques and also introduce the problem of document sanitization.

Martínez et al. [20] (Chap. 7) consider the anonymization of categorical datasets using semantic information. The authors consider well known anonymization methods from SDC such as recoding, microaggregation, and resampling. These methods are then adapted to take into account the semantics of the data they are protecting, usually relaying in ontologies to model the semantic knowledge associated with the attributes of the dataset.

Batet and Sanchez [5] (Chap. 8) go in depth with semantic privacy methods by reviewing semantic similarity functions. Several SDC and PPDM methods such as microaggregation, additive noise, recoding, sampling, or data swapping, require to some extent the use of a distance or similarity function. The chapter serves as a survey of semantic similarity functions to be used in such privacy protection methods.

Nettleton and Abril [28] (Chap. 9) tackle the problem of document sanitization. The sanitization process allows to disclose a confidential document by removing, generalizing, or distorting the confidential information contained in the document. The authors evaluate the sanitization process using information retrieval metrics.

4.3 Location Privacy

A very specific type of data that has gained recent popularity is the one related to localization. The advances in localization technology have made it very easy to collect location data from smartphones, GPS devices,.... The possibility of mining these data opens up interesting application, but at the same time expose the privacy of the respondents of the data. Here we will see three approaches to deal with location data, two of them treat trajectory data, which consider location and timing (e.g. vehicle trajectories within a urban environment).

Conti et al. [8] (Chap. 10) review user privacy in location based services based on footprints. A footprint considers the amount of time that the user spends in a given area. The authors show the risk and weakness found in this type of anonymization models when facing an adversary with previous knowledge not considered by the anonymization procedure. This analysis leads the authors to conclude with a set of properties to determine the actual level of privacy of these models. In this scenarios the actual anonymization is not performed by the users as active subjects but by a trusted server, the so called location depersonalization server. It is this service who also discloses the protected data regarding the users.

Trujillo-Rasua et al. [38] (Chap. 11) depict a review of privacy methods for spatio-temporal databases. More precisely, authors provide a review of microaggregation to protect data related to movement, or trajectories. A trajectory can be seen as a location data with a temporal component. The chapter is concluded with a concrete proposal and evaluation of an specific microaggregation method for trajectories.

Martínez-Bea [22] (Chap. 12) also considers the anonymization of data describing trajectories using microaggregation. In this case, the protection mechanism is based on time series. That is, the proposal departs from previous work on the anonymization of time series and applies it to trajectories. As the chapter describes, this method was implemented as part of a demonstrator of the ARES project [3].

5 Social Networks

Social networks, by their intrinsic nature, expose sensitive information from their users. Protecting the privacy of users in social networks is a hot research topic. When we consider the respondent privacy approaches in social networks, we are assuming that the network authority or a trusted third party performs the anonymization. In this case we will also see metrics to measure privacy.

Casas-Roma et al. [7] (Chap. 13) consider the protection of graph data. These data usually corresponds to social network relationships, which can be considered as sensitive information. The authors review privacy preserving methods for networks based on the k -anonymity property. The chapter includes an empirical evaluation of the methods.

Sramka [33] (Chap. 14) provides a review of privacy metrics in the context of social networks. Furthermore, the author introduces a novel privacy metric. These metrics are very useful in order to assess the privacy exposure of the users of a social network. Users can be aware of how their sensitive information is being distributed in the network. We have classified this work as respondent privacy since the computation

of the proposed metric requires the consideration of the whole network. Something that, in some cases, is only available to the social network administrative authority and not individual users.

5.1 Other Respondent Privacy Enhancing Technologies

The consideration of privacy in other systems to protect the anonymity of the respondents is also increasing in recent years. Privacy is being considered in authentication schemes, electronic ticketing systems and coupons booklets, in RFID technology, or in video surveillance.

Martínez-Ballester et al. [21] (Chap. 15) present a review of Trustworthy Video Surveillance System (T-VSS). Their work faces the problem of anonymizing surveillance video files to mitigate the disclosure of individual personal data. The authors present a privacy-aware video surveillance platform that can be used as a safety protection while preserving the privacy of individuals.

Payeras-Capellà et al. [30] (Chap. 16) introduce the study of privacy issues in electronic ticketing systems. They analyze the requirements and state of the art in electronic tickets used in transport services, and highlight required properties regarding the privacy and anonymity of users.

Garcia-Alfaro et al. [14] (Chap. 17) consider privacy issues related to passive RFID tags. More precisely the authors introduce and analyze the EPC Gen2 technology identifying security and privacy threats. The authors also survey countermeasures applicable to mitigate the identified threats. The chapter also outlines the work done within the ARES project in relation to this topic and discuss future research directions.

Hinarejos et al. [15] (Chap. 18) consider electronic coupons booklets. The authors review the state of the art of these systems which are the electronic equivalent of paper coupons booklets, usually offered as discount tickets to users. The chapter outlines the security and privacy requirements of these systems, and propose a solution for mobile scenarios.

Daza and Signorini [9] (Chap. 19) review authentication technologies taking into special consideration its use as a privacy enhancing technology. Moreover the authors discuss hardware intrinsic security (HIS) approaches and present the APtItUDE system which while using physically unclonable functions, avoids the use of challenge-response databases. This system guarantees a high level of data privacy while providing a user friendly authentication process.

6 User Privacy

This other part of the book deals with the user privacy approaches. We will see several systems and methods where the users perform active actions to ensure the protection of their privacy. As an example of user privacy we will consider how users can actively protect their privacy against a web search engine, or recommender and personalized information systems.

6.1 Web Search Engines

The information gathered by a Web search engine (WSE) can undoubtedly raise important privacy concerns to their users. There are two approaches to allow users to use a WSE without revealing their sensitive information. The first one is when the users trust the WSE to perform an anonymization procedure on the gathered data [27, 31]. This will be the case of a respondent privacy approach to protect WSE information. The second approach, on the contrary, relies on the users themselves to perform the required actions to ensure their own privacy. We will see here examples of this second approach.

Romero-Tris et al. [32] (Chap. 20) review the so called multi-party approach for user anonymization of queries. These kind of methods rely in multi-party protocols performed by a group of cooperative users in order to hide the real preferences of each individual user within the group. The authors also propose some improvements over the Useless User Profile protocol, which allow users to security exchange their queries. This allows each user to submit a query from her partners distorting the profile that the WSE can build for her.

Juarez and Torra [17] (Chap. 21) also deal with the problem of anonymizing user profiles to a WSE. They discuss DisPA (Dissociating Privacy Agent), a browser extension, which allows the user to increase its privacy against a WSE. To do that, DisPA semantically disassociates search queries by topics, which are sent with different profiles. To the eyes of the WSE these are queries coming from different users, but given that they are semantically grouped, they allow certain degree of personalization by the WSE.

Stokes and Bras-Amorós [34] (Chap. 22) introduce the use of combinatorial configurations to model peer-to-peer private information retrieval protocols. Although it refers to UPIR protocols, it can be contextualized also in terms of WSE, and more precisely as multi-party methods from [32] or the untrusted model discussed in [29]. The users collaborate to perform the query in a database. The authors provide a review and introduce important concepts and approaches to deal specifically with user collusion and anonymous neighbors.

6.2 Recommender and Personalized Systems

Recommender and personalized systems are becoming very important. Contrary to what we have seen regarding social networks (see Sect. 5), here we will show user privacy methods in a very similar context. In these cases it is the user who, to some extent, takes action to protect its own sensitive information.

Parra-Arnau et al. [29] (Chap. 23) provide a review of privacy in personalized information systems. The authors review the state of the art and classify existing proposals. At the same time they also review privacy metrics for this personalized information systems. In their review, the authors distinguish between a *trusted* model, which requires a trusted third party to perform the anonymization, *untrusted* model, where there is no trusted party and the anonymization procedure relies in the users, and *semi-trusted* model, when the users collaborate among peers to perform the anonymization (in the same line as the multi-party methods for queries presented in [32]). Regarding our broad classification, the trusted model will yield respondent privacy methods, while untrusted and semi-trusted models correspond to user privacy.

We have included this work as a user privacy approach since we feel it can be more interesting from this point of view.

Vera del Campo et al. [39] (Chap. 24) address the problem of privacy in recommendation systems. The work is contextualized in the Internet of Things, and presents DocCloud. DocCloud is a document recommender system, which provides several privacy-related protections, which here is extended to generic cloud resources in the context of a social network. We have classified this work as user privacy, although the anonymization is somehow ensured by the infrastructure.

7 Conclusions

This chapter introduces the current book on data privacy. Although this is not an exhaustive enumeration, the book gives a broad picture of the work done under the umbrella of the ARES research project.

The chapters of the book are contextualized in respondent and user privacy. The chapters present state of the art research in several fields such as statistical disclosure control and privacy-preserving data mining, security technologies, and social networks.

Acknowledgments

Partial support by the Spanish MICINN (projects COPRIVACY (TIN2011-27076-C03-03), N-KHRONOUS (TIN2010-15764), and ARES (CONSOLIDER INGENIO 2010 CSD2007-00004)) and by the EC (FP7/2007-2013) Data without Boundaries (grant agreement number 262608) is acknowledged.

References

1. Abril, D., Navarro-Arribas, G., Torra, V.: Data privacy with R. Chapter 17, *Advanced Research on Data Privacy*. Springer, Cham (2014)
2. Aggarwal, C.C., Yu, P.S.: A general survey of privacy-preserving data mining models and algorithms. *Privacy-Preserving Data Mining, Advances in Database Systems*, pp. 11–52. Springer, New York (2008)
3. Aragonés, J., Manjón, J.A.: Field trial for joint validation and media dissemination of WP1-WP2-WP3-WP4 technologies in a real-world vehicular network environment (WP6.T1) Deliverable Report. ARES project CONSOLIDER-INGENIO 2010 CSD2007-00004 (2012)
4. Barbaro, M., Zeller, T.: A Face is Exposed for AOL Searcher No. 4417749. *The New York Times*, New York (2006). Accessed 9 Aug 2006 (Accessed 25 Apr 2010)
5. Batet, M., Sanchez, D.: Contributions on Semantic Similarity and its Applications to Data Privacy. Chapter 18, *Advanced Research on Data Privacy*. Springer, Cham (2014)
6. BBC News: Sony faces legal action over attack on PlayStation network. BBC news technology. <http://www.bbc.co.uk/news/technology-13192359> (2011). Accessed 28 Apr 2011
7. Casas-Roma, J., Herrera-Joancomartí, J., Torra, V.: A Summary of k-Degree Anonymous Methods for Privacy-Preserving on Networks. Chapter 13, *Advanced Research on Data Privacy*. Springer, Cham (2014)
8. Conti, M., Di Pietro, R., Marconi, L.: Privacy for LBSs: on Using a Footprint Model to Face the Enemy. Chapter 10, *Advanced Research on Data Privacy*. Springer, Cham (2014)