# Advances in Information Security

## Volume 57

*Series Editor*

Sushil Jajodia, Center for Secure Information Systems, George Mason University, Fairfax, VA, 22030-4444, USA

More information about this series at http://www.springer.com/series/5576

Giovanni Livraga

# Protecting Privacy in Data Release

Springer

Giovanni Livraga
Universita degli Studi di Milano
Crema, Italy

ISSN 1568-2633
Advances in Information Security
ISBN 978-3-319-16108-2 ISBN 978-3-319-16109-9 (eBook)
DOI 10.1007/978-3-319-16109-9

Library of Congress Control Number: 2015935559

Springer Cham Heidelberg New York Dordrecht London


Printed on acid-free paper

Springer International Publishing AG Switzerland is part of Springer Science+Business Media (www.springer.com)

*To my family*
*To Cesare*

# Preface

Data sharing and dissemination play a key role in our information society. Not only do they prove to be advantageous to the involved parties, but they can also be fruitful to the society at large: for instance, new treatments for rare diseases can be discovered with real clinical trials shared by hospitals and pharmaceutical companies. The advancements in the Information and Communication Technology (ICT) make the process of releasing a data collection simpler than ever. The availability of novel computing paradigms, such as data outsourcing and cloud computing, makes scalable, reliable, and fast infrastructures a dream come true at reasonable costs. As a natural consequence of this scenario, data owners often rely on external servers for releasing their data collections, thus delegating the burden of data storage and management to the service provider. Unfortunately, the price to be paid is in terms of unprecedented privacy and security risks. Data collections often include sensitive information, not intended for disclosure, that should be properly protected. The problem of protecting privacy in data release has been under the attention of the research and development communities for a long time. However, the richness of released data, the large number of available sources, and the emerging outsourcing/cloud scenarios raise novel problems, not addressed by traditional approaches, which call for enhanced solutions.

In this book, we propose a comprehensive approach for protecting sensitive information when large collections of data are publicly or selectively released by their owners. In a nutshell, this requires protecting data explicitly included in the release, as well as protecting information not explicitly released but that could be exposed by the release, and ensuring that access to released data be allowed only to authorized parties according to the data owners' policies. More specifically, these three aspects translate to three requirements, addressed by this book, which can be summarized as follows. The first requirement is the *protection of data explicitly included in a release*. While intuitive, this requirement is complicated by the fact that privacy-enhancing techniques should not prevent recipients from performing legitimate analysis on the released data but, on the contrary, should ensure sufficient visibility over non sensitive information. We therefore propose a solution, based on a novel formulation of the fragmentation approach, that vertically fragments

a data collection so to satisfy requirements for both information protection and visibility, and we complement it with an effective technique for enriching the utility of the released data. The second requirement is the *protection of data not explicitly included in a release*. As a matter of fact, even a collection of non sensitive data might enable recipients to infer (possibly sensitive) information not explicitly disclosed but that somehow depends on the released information (e.g., the release of the treatment with which a patient is being cared can leak information about his/her disease). To address this requirement, starting from a real case study, we propose a solution for counteracting the inference of sensitive information that can be drawn observing peculiar value distributions in a released data collection. The third requirement is *access control enforcement*. Available solutions fall short in emerging computing paradigms for a variety of reasons. Traditional access control mechanisms are in fact typically based on a reference monitor mediating access requests, and do not fit outsourcing/cloud scenarios where neither data owners are willing nor cloud providers are trusted, to enforce the authorization policy. Recent solutions applicable to outsourcing scenarios assume outsourced data to be read-only and cannot easily manage (dynamic) write authorizations. We therefore propose an approach for efficiently supporting grant and revoke of write authorizations, building upon the selective encryption approach, and we also define a subscription-based authorization policy, to fit real-world scenarios where users pay for a service and access the resources made available during their subscriptions.

The main contributions of this book can therefore be summarized as follows.

- With respect to the protection of data explicitly included in a release, our original results are: (1) a novel modeling of the fragmentation problem; (2) an efficient technique for computing a fragmentation, based on reduced Ordered Binary Decision Diagrams (OBDDs) to formulate the conditions that a fragmentation must satisfy; (3) the computation of a minimal fragmentation not fragmenting data more than necessary, with the definition of both an exact and a heuristics algorithm providing faster computational time while well approximating the exact solutions; (4) the definition of loose associations, a sanitized form of the sensitive associations broken by fragmentation, specifically designed to operate on arbitrary fragmentations; and (5) the definition of a heuristic algorithm for the computation of arbitrary loose associations, experimentally proved to enhance precision of queries executed over different fragments.
- With respect to the protection of data not explicitly included in a release, our original results are: (1) the definition of a novel and unresolved inference scenario, raised from a real case study where data items are incrementally released upon request; (2) the definition of several metrics to assess the inference exposure due to a data release, based upon the concepts of mutual information, Kullback–Leibler distance between distributions, Pearson's cumulative statistic, and Dixon's coefficient; (3) the identification of a safe release with respect to a given inference channel; and (4) the definition of the controls to be enforced to guarantee that no sensitive information be leaked releasing non sensitive data items.

- With respect to the access control enforcement, our original results are: (1) the management of dynamic write authorizations, by defining a solution based on selective encryption for efficiently and effectively supporting grant and revoke of write authorizations; (2) the definition of an effective technique to guarantee data integrity, so to allow the data owner and the users to verify that modifications to a resource have been produced only by authorized users; and (3) the modeling and enforcement of a subscription-based authorization policy, to support scenarios where both the set of users and the set of resources change frequently over time, and users' authorizations to access resources are based on their subscriptions.

Milan, Italy Giovanni Livraga
November 2014

# Acknowledgements

This book is the result of the publication of my Ph.D. thesis, and there is a long list of people I am indebted to and to which I want to express my gratitude.

I must first sincerely thank my advisor Prof. Pierangela Samarati, for having dedicated her time to follow me in my work with constant presence and valuable and fruitful supervision. I am particularly grateful for the many opportunities I got from her, among which the possibility of preparing this book is just a little example.

I am grateful to Prof. Sabrina De Capitani di Vimercati, for her valuable guidance, profitable comments on the work, and optimistic spin I have always got from her, which helped to find the right track in many occasions.

I am also indebted to Dr. Sara Foresti, for always having her office door open when I knock, and being ready to provide help and many corrections to my work.

I would like to express my gratitude to Prof. Sushil Jajodia, who first envisioned the opportunity to publish this book based on my Ph.D. thesis. I am grateful to him also for the opportunity to visit the Center for Secure Information Systems at George Mason University, USA, where I could find a stimulating and pleasant environment.

I thank Dr. Michele Bezzi, Dr. Valentina Ciriani, Prof. Stefano Paraboschi, and Dr. Roberto Sassi, for the valuable discussions and their support on different aspects of the work I present in this book. I would like to thank Prof. Vijay Atluri, Prof. Sushil Jajodia, and Prof. Javier Lopez, for providing precious comments and suggestions that improved the presentation of this work.

My gratitude goes also to Susan Lagerstrom-Fife and Jennifer Malat, for their support in the preparation of this book.

A special mention to my family: mum and dad, for bringing me up the way I am and for being two examples that have motivated me to reach this goal; Chiara, for the lovely chats on Tuesday lunch breaks; Matteo, for the help I can have when I need it.

Cesare: Your support is and has been fundamental to me, also in reaching this goal. You are always right next to me when I need it, and this is just priceless to me. And thanks, for patiently caring about me and, more patiently, teaching me to care about myself. Having you by my side makes me a lucky man.

In random order, The Brooke's inner circle: Gemma, Plitz, Soult, Vitti, Ilaria, Marta and Claudio. Riki and Stefi. Alice and Michele and our coffee breaks. Ruggero, Angelo, Paolo, Gerson, and the other (present and past) Ph.D. students and colleagues at the department. Chiara and Nora, wonderful *quasi*-flatmates, and helpful friends in troublesome occasions. Federico, Francesca, Erika, and all the other guys, for having opened their Gran Sasso door to me.

A last, special, though to my grandmothers Augusta and Alice who, I am sure, are the most proud of all.

# Contents