Medical Data Privacy Handbook

Aris Gkoulalas-Divanis • Grigorios Loukides
Editors

Medical Data Privacy Handbook

Springer

Editors
Aris Gkoulalas-Divanis
IBM Research - Ireland
Mulhuddart
Dublin, Ireland

Grigorios Loukides
Cardiff University
Cardiff, UK

ISBN 978-3-319-23632-2 ISBN 978-3-319-23633-9 (eBook)
DOI 10.1007/978-3-319-23633-9

Library of Congress Control Number: 2015947266

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)

Dedicated to my parents, to Elena, and to the memory of my grandmother Sophia

–Aris Gkoulalas-Divanis

Dedicated to the memory of my grandmother

–Grigorios Loukides

Preface

The editors started working on medical data privacy in 2009, when they were postdoctoral researchers in the Health Information Privacy Laboratory, Department of Biomedical Informatics, Vanderbilt University. Their work on the topic involved understanding the privacy risks of medical data publishing and developing methods to prevent these risks. Protecting medical data privacy is a challenging problem, since a large volume of complex data must be protected in a setting that involves multiple parties (patients, physicians, carers, researchers, etc.). To address the problem, it is important to develop principled approaches that are specifically geared towards medical data. In addition, it is equally important to increase the awareness of all parties, involved in managing medical data, about privacy risks and approaches for achieving medical data privacy. Thus, the overarching aim of this book is to survey the field of medical data privacy and to present the state-of-the-art approaches to a wide audience.

The structure of the book closely follows the main categories of research works that have been undertaken to protect medical data privacy. Each such category is surveyed in a different part of the book, as follows. Part I is devoted to medical data *sharing*. Part II focuses on medical data privacy in *distributed and dynamic settings*. Following that, Part III examines privacy preservation in *emerging applications* featuring medical data, and Part IV discusses medical data privacy through *policy*, *data de-identification*, and *data governance*.

Privacy-preserving data sharing requires protecting the identity of patients and/or their sensitive information. For instance, attackers may use external data or background knowledge to learn patients' identity, even though attributes that directly identify patients (e.g., SSNs, phone numbers) have been removed. The problem has been studied extensively in the context of medical data, by the computer science, medical informatics, and statistics communities. However, there is no one-size-fits-all solution and various challenges remain. The purpose of Part I of this book is to survey the main research directions in the area of privacy-preserving medical data sharing and to present state-of-the-art approaches, including measures, algorithms, and software tools, that have been designed to solve this problem.

The protection of medical data privacy is particularly challenging, when multiple interrelated parties are involved. For example, medical data practitioners often need to link or exchange different parts of data about a patient, in the context of patient treatment. In addition, medical researchers or insurers may need to access patient information, according to the patient's privacy requirements. In this case, both the objectives of the parties accessing the data and the patient's requirements may change over time. Furthermore, data that are stored or processed in the cloud are vulnerable to a multitude of attacks, ranging from malicious access to intentional data modification. Part II of this book presents approaches focusing on privacy protection in such distributed and dynamic settings. These include approaches for linking data (record linkage), managing data access and patient consent, as well as exchanging health information. Furthermore, a comprehensive survey of privacy concerns and mitigation strategies for medical data in the cloud is presented.

Advances in medical devices and ubiquitous computing enable the collection and analysis of many complex data types, including genomic data, medical images, sensor data, biomedical signals, and health social network data. These data are valuable in a wide spectrum of emerging applications, either alone or in combination with data such as patient demographics and diagnosis codes, which are commonly found in Electronic Health Record (EHR) systems. For example, genomic studies have strong potential to lead to the discovery of effective, personalized drugs, and therapies. However, genomic data are extremely sensitive and must be privacy-protected. Part III of this book surveys privacy threats and solutions for all the aforementioned types of data that are central in emerging applications.

Parts I–III of this book focus on technical solutions that allow data owners (e.g., a healthcare institution) to effectively protect medical data privacy. On the other hand, Part IV focuses on the legal requirements for offering data privacy protection, as well as on the techniques and procedures that are required to satisfy this requirement. More specifically, this part examines key legal frameworks related to medical data privacy protection, as well as data de-identification and governance solutions, which are required to comply with these frameworks. A detailed presentation of the data protection legislation in the USA, EU, UK, and Canada is offered.

This book is primarily addressed to researchers and educators in the areas of computer science, statistics, and medical informatics who are interested in topics related to medical privacy. This book will also be a valuable resource to industry developers, as it explains the state-of-the-art algorithms for offering privacy. To ease understanding by nonexperts, the chapters contain a lot of background material, as well as many examples and citations to related literature. In addition, knowledge of medical informatics methods and terminology is not a prerequisite, and formalism was intentionally kept at a minimum. By discussing a wide range of privacy techniques, providing in-depth coverage of the most important ones, and highlighting promising avenues for future research, this book also aims at attracting computer science and medical informatics students to this interesting field of research.

Dublin, Ireland Aris Gkoulalas-Divanis
Cardiff, UK Grigorios Loukides
July, 2015

Acknowledgements

We would like to thank all the authors, who have contributed chapters to this book, for their valuable contributions. This work would not have been possible without their efforts. A total of 63 authors who hold positions in leading academic institutions and industry, in Europe (France, Germany, Greece, Italy, Luxembourg, Switzerland, and UK), North America, Asia, Australia, and New Zealand, have contributed 29 chapters in this book, featuring more than 280 illustrations. We sincerely thank them for their hard work and the time they devoted to this effort.

In addition, we would like to express our deep gratitude to all the expert reviewers of the chapters for their constructive comments, which significantly helped towards improving the organization, readability, and overall quality of this handbook.

Last but not least, we are indebted to Susan Lagerstrom-Fife and Jennifer Malat from Springer, for their great guidance and advice in the preparation and completion of this handbook, as well as to the publication team at Springer for their valuable assistance in the editing process.

Contents

List of Figures

List of Tables