

Advances in Information Security

Volume 69

Series editor

Sushil Jajodia, George Mason University, Fairfax, VA, USA

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

Tianqing Zhu • Gang Li • Wanlei Zhou
Philip S. Yu

Differential Privacy and Applications

 Springer

Tianqing Zhu
Deakin University
Melbourne, Australia

Gang Li 
Deakin University
Melbourne, Australia

Wanlei Zhou
Deakin University
Melbourne, Australia

Philip S. Yu
University of Illinois at Chicago
Chicago, IL, USA

ISSN 1568-2633

Advances in Information Security

ISBN 978-3-319-62002-2

ISBN 978-3-319-62004-6 (eBook)

DOI 10.1007/978-3-319-62004-6

Library of Congress Control Number: 2017946488

© Springer International Publishing AG 2017

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.

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.

The publisher, the authors and the editors are safe to assume that the advice and information in this book are believed to be true and accurate at the date of publication. Neither the publisher nor the authors or the editors give a warranty, express or implied, with respect to the material contained herein or for any errors or omissions that may have been made. The publisher remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG

The registered company address is: Gewerbestrasse 11, 6330 Cham, Switzerland

Preface

Corporations, organizations, and governments have collected, digitized, and stored information in digital forms since the invention of computers, and the speed of such data collection and volumes of stored data have increased dramatically over the past few years, thanks to the pervasiveness of computing devices and the associated applications that are closely linked to our daily lives. For example, hospitals collect records of patients, search engines collect online behaviors of users, social network sites collect connected friends of people, e-commerce sites collect shopping habits of customers, and toll road authorities collect travel details of vehicles. Such huge amounts of datasets provide excellent opportunities for businesses and governments to improve their services and to bring economic and social benefits, especially through the use of technologies dealing with big data, including data mining, machine learning, artificial intelligence, data visualization, and data analytics. For example, by releasing some statistics of hospital records may help medical research to combat diseases. However, as most of the collected datasets are personally related and contain private or sensitive information, data releasing may also provide a fertile ground for adversaries to obtain certain private and sensitive information, even though simple anonymization techniques are used to hide such information. Privacy preserving has, therefore, become an urgent issue that needs to be addressed in the digital age.

Differential privacy is a new and promising privacy framework and has become a popular research topic in both academia and industry. It is one of the most prevalent privacy models as it provides a rigorous and provable privacy notion that can be applied in various research areas, and can be potentially implemented in various application scenarios. The goal of this book is to summarize and analyze the state-of-the-art research and investigations in the field of differential privacy and its applications in privacy-preserving data publishing and releasing, so as to provide an approachable strategy for researchers and engineers to implement this new framework in real world applications.

This is the first book with a balanced view on differential privacy theory and its applications, as most existing books related to privacy preserving either do not

touch the topic of differential privacy or only focus on the theoretical analysis of differential privacy. Instead of using abstract and complex notions to describe the concepts, methods, algorithms, and analysis on differential privacy, this book presents these difficult topics in a combination of applications, in order to help students, researchers, and engineers with less mathematical background understand the new concepts and framework, enabling a wider adoption and implementation of differential privacy in the real world. The striking features of the book, differs from others, can be illustrated from three basic aspects:

- A detailed coverage on differential privacy in the perspective of engineering, rather than computing theory. The most difficult part in comprehending differential privacy is the complexity and the level of abstract of the theory. This book presents the theory of differential privacy in a more natural and easy to understand way.
- A rich set of state-of-the-art examples on various application areas helps readers to understand how to implement differential privacy in real world scenarios. Each application example includes a brief introduction to the problem and its challenges, a detailed implementation of differential privacy to solve the problem, and an analysis on the privacy and utility.
- A comprehensive collection of contemporary research results and issues in differential privacy. Apart from the basic theory, most of the contents of the book are from the recent publications in the last 5 years, reflecting the state-of-the-art of research and development in the area of differential privacy.

This book intends to enable readers, especially postgraduate and senior undergraduate students, to study up-to-date concepts, methods, algorithms, and analytic skills for building modern privacy-preserving applications through differential privacy. It enables students not only to master the concepts and theories in relation to differential privacy but also to readily use the material introduced into implementation practices. Therefore, the book is divided into two parts: theory and applications. In the theory part, after an introduction of the differential privacy preliminaries, the book presents detailed descriptions from an engineering viewpoint on areas of differentially private data publishing and differentially private data analysis where research on differential privacy has been conducted. In the applications part, after a summary on the steps to follow when solving the privacy-preserving problem in a particular application, the book then presents a number of state-of-the-art application areas of differential privacy, including differentially private social network data publishing, differentially private recommender system, differential location privacy, spatial crowdsourcing with differential privacy preservation, and correlated differential privacy for non-IID datasets. The book also includes a final chapter on the future direction of differential privacy and its applications.

Acknowledgments

We are grateful to many research students and colleagues at Deakin University in Melbourne and University of Illinois at Chicago, who have made a lot of comments to our presentations as their comments inspire us to write this book. We would like to acknowledge some support from research grants we have received, in particular, the Australian Research Council Grant no. DP1095498, LP120200266, and DP140103649, NSF through grants IIS-1526499, and CNS-1626432, and NSFC (Nos. 61672313, 61502362). Some interesting research results presented in the book are taken from our research papers that indeed (partially) supported through these grants. We also would like to express our appreciations to the editors in Springer, especially Susan Lagerstrom-Fife, for the excellent professional support. Finally we are grateful to the family of each of us for their consistent and persistent supports. Without their support, the book may just become some unpublished discussions.

Melbourne, Australia
Melbourne, Australia
Melbourne, Australia
Chicago, IL, USA
May 2017

Tianqing Zhu
Wanlei Zhou
Gang Li
Philip S. Yu

Contents

1	Introduction	1
1.1	Privacy Preserving Data Publishing and Analysis	1
1.2	Privacy Violations	2
1.3	Privacy Models	3
1.4	Differential Privacy	4
1.5	Outline and Book Overview	5
2	Preliminary of Differential Privacy	7
2.1	Notations	7
2.2	Differential Privacy Definition	9
2.2.1	The Privacy Budget	9
2.3	The Sensitivity	10
2.3.1	The Global Sensitivity	11
2.3.2	The Local Sensitivity	11
2.4	The Principle Differential Privacy Mechanisms	12
2.4.1	The Laplace Mechanism	13
2.4.2	The Exponential Mechanism	14
2.5	Utility Measurement of Differential Privacy	15
3	Differentially Private Data Publishing: Settings and Mechanisms	17
3.1	Interactive and Non-interactive Settings	17
3.2	Publishing Mechanism	19
4	Differentially Private Data Publishing: Interactive Setting	23
4.1	Transaction Data Publishing	23
4.1.1	Laplace	23
4.1.2	Transformation	24
4.1.3	Query Separation	24
4.1.4	Iteration	25
4.1.5	Discussion	25

4.2	Histogram Publishing	26
4.2.1	Laplace	27
4.2.2	Partition of Dataset	27
4.2.3	Consistency of Histogram	28
4.3	Stream Data Publishing	29
4.3.1	Laplace	30
4.3.2	Partition of Dataset	30
4.3.3	Iteration	30
4.3.4	Discussion	31
4.4	Graph Data Publishing	31
4.4.1	Edge Differential Privacy	32
4.4.2	Node Differential Privacy	33
4.4.3	Discussion	34
4.5	Summary on Interactive Setting	34
5	Differentially Private Data Publishing: Non-interactive Setting	35
5.1	Batch Queries Publishing	35
5.1.1	Laplace	36
5.1.2	Transformation.....	36
5.1.3	Partition of Dataset	38
5.1.4	Iteration	38
5.1.5	Discussion	38
5.2	Contingency Table Publishing	39
5.2.1	Laplace	39
5.2.2	Iteration	40
5.2.3	Transformation.....	40
5.3	Anonymized Dataset Publishing	41
5.4	Synthetic Dataset Publishing	43
5.4.1	Synthetic Dataset Publishing Based on Learning Theory.....	43
5.4.2	High Dimensional Synthetic Dataset Publishing.....	47
5.5	Summary on Non-interactive Setting	48
6	Differentially Private Data Analysis	49
6.1	Laplace/Exponential Framework	49
6.1.1	SuLQ and PINQ Interface	50
6.1.2	Specific Algorithms in the Laplace/Exponential Framework	51
6.1.3	Summary on Laplace/Exponential Framework	57
6.2	Private Learning Framework.....	57
6.2.1	Foundation of ERM	58
6.2.2	Private Learning in ERM.....	59
6.2.3	Sample Complexity Analysis	62
6.2.4	Summary on Private Learning Framework	64
6.3	Summary of Differentially Private Data Analysis	65

- 7 Differentially Private Deep Learning** 67
 - 7.1 Introduction 67
 - 7.2 Preliminary 69
 - 7.2.1 Deep Learning Structure 69
 - 7.2.2 Stochastic Gradient Descent 71
 - 7.3 Differentially Private Deep Learning 73
 - 7.3.1 Basic Laplace Method 74
 - 7.3.2 Private SGD Method 75
 - 7.3.3 Deep Private Auto-Encoder Method 77
 - 7.3.4 Distributed Private SGD 79
 - 7.4 Experimental Methods 81
 - 7.4.1 Benchmark Datasets 81
 - 7.4.2 Learning Objectives 81
 - 7.4.3 Computing Frameworks 82
 - 7.5 Summary 82
- 8 Differentially Private Applications: Where to Start?** 83
 - 8.1 Solving a Privacy Problem in an Application 83
 - 8.2 Challenges in Differentially Private Applications 85
 - 8.2.1 High Sensitivity Challenge 85
 - 8.2.2 Dataset Sparsity Challenge 85
 - 8.2.3 Large Query Set Challenge 86
 - 8.2.4 Correlated Data Challenge 86
 - 8.2.5 Computational Complexity Challenge 87
 - 8.2.6 Summary 87
 - 8.3 Useful Public Datasets in Applications 88
 - 8.3.1 Recommender System Datasets 88
 - 8.3.2 Online Social Network Datasets 89
 - 8.3.3 Location Based Datasets 89
 - 8.3.4 Other Datasets 89
 - 8.4 Applications Settings 90
- 9 Differentially Private Social Network Data Publishing** 91
 - 9.1 Introduction 91
 - 9.2 Preliminaries 92
 - 9.3 Basic Differentially Private Social Network Data Publishing Methods 93
 - 9.3.1 Node Differential Privacy 93
 - 9.3.2 Edge Differential Privacy 97
 - 9.4 Graph Update Method 98
 - 9.4.1 Overview of Graph Update 98
 - 9.4.2 Graph Update Method 100
 - 9.4.3 Update Function 101
 - 9.4.4 Privacy and Utility Analysis 101
 - 9.4.5 Experimental Evaluation 103
 - 9.5 Summary 105

10	Differentially Private Recommender System	107
10.1	Introduction	107
10.2	Preliminaries	109
10.2.1	Collaborative Filtering	109
10.2.2	Neighborhood-Based Methods: k Nearest Neighbors	109
10.2.3	Model-Based Methods: Matrix Factorization	111
10.3	Basic Differentially Private Recommender Systems	112
10.3.1	Differentially Private Untrustworthy Recommender System	113
10.3.2	Differentially Private Trustworthy Recommender System	114
10.4	Private Neighborhood-Based Collaborative Filtering Method	117
10.4.1	KNN Attack to Collaborative Filtering	117
10.4.2	The Private Neighbor Collaborative Filtering Algorithm	118
10.4.3	Privacy and Utility Analysis	123
10.4.4	Experiment Analysis	126
10.5	Summary	129
11	Privacy Preserving for Tagging Recommender Systems	131
11.1	Introduction	131
11.2	Preliminaries	133
11.2.1	Notations	133
11.2.2	Tagging Recommender Systems	133
11.2.3	Related Work	134
11.3	Private Tagging Publishing Method	134
11.3.1	User Profiles	134
11.3.2	Private Tagging Release Algorithm Overview	136
11.3.3	Private Topic Model Generation	137
11.3.4	Topic Weight Perturbation	139
11.3.5	Private Tag Selection	141
11.3.6	Privacy and Utility Analysis	143
11.3.7	Experimental Evaluation	146
11.4	Summary	149
12	Differentially Location Privacy	151
12.1	Introduction	151
12.2	Preliminary	152
12.3	Basic Location Privacy Methods	153
12.3.1	Snapshot Location Privacy: Geo-Indistinguishability	154
12.3.2	Trajectory Privacy	157
12.4	Hierarchical Snapshot Location Publishing	160
12.4.1	Hierarchical Sensitivity	160
12.4.2	Overview of Private Location Release	162
12.4.3	Private Location Release Algorithm	163
12.4.4	Utility and Privacy	167
12.4.5	Experimental Evaluation	170
12.5	Summary	172

- 13 Differentially Private Spatial Crowdsourcing** 173
 - 13.1 Introduction 173
 - 13.2 Basic Method 174
 - 13.2.1 Background of Crowdsourcing 174
 - 13.2.2 Differentially Private Crowdsourcing Methods 175
 - 13.3 Differential Privacy in Reward-Based Crowdsourcing 177
 - 13.3.1 Problem Statement 178
 - 13.3.2 Building a Contour Plot with DP Guarantee 178
 - 13.3.3 Task Assignment..... 181
 - 13.3.4 Experimental Evaluation 186
 - 13.4 Summary..... 189
- 14 Correlated Differential Privacy for Non-IID Datasets**..... 191
 - 14.1 Introduction..... 191
 - 14.2 An Example: Correlated Records in a Dataset 192
 - 14.3 Basic Methods 194
 - 14.3.1 Pufferfish 194
 - 14.3.2 Blowfish 195
 - 14.4 Correlated Differential Privacy 196
 - 14.4.1 Correlated Differential Privacy Problem 196
 - 14.4.2 Research Issues and Challenges 197
 - 14.4.3 Correlated Dataset Analysis..... 198
 - 14.4.4 Correlated Sensitivity 199
 - 14.4.5 Correlated Iteration Mechanism 201
 - 14.4.6 Mechanism Analysis 206
 - 14.4.7 Experiment and Analysis..... 208
 - 14.5 Summary..... 214
- 15 Future Directions and Conclusion** 215
 - 15.1 Adaptive Data Analysis: Generalization in Machine Learning 215
 - 15.2 Personalized Privacy 216
 - 15.3 Secure Multiparty Computations with Differential Privacy 216
 - 15.4 Differential Privacy and Mechanism Design 217
 - 15.5 Differential Privacy in Genetic Data 217
 - 15.6 Local Differential Privacy..... 218
 - 15.7 Learning Model Publishing 220
 - 15.8 Conclusion 222
- References**..... 223
- Index**..... 235