

Wen Ming Liu
Lingyu Wang

Preserving Privacy Against Side-Channel Leaks

From Data Publishing to Web
Applications

Advances in Information Security

Volume 68

Series Editor

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

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

Wen Ming Liu • Lingyu Wang

Preserving Privacy Against Side-Channel Leaks

From Data Publishing to Web Applications

 Springer

Wen Ming Liu
Concordia Institute for Information
Systems Engineering
Concordia University
Montreal, QC, Canada

Lingyu Wang
Concordia Institute for Information
Systems Engineering
Concordia University
Montreal, QC, Canada

ISSN 1568-2633

Advances in Information Security

ISBN 978-3-319-42642-6

ISBN 978-3-319-42644-0 (eBook)

DOI 10.1007/978-3-319-42644-0

Library of Congress Control Number: 2016948836

© Springer International Publishing Switzerland 2016

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.

Printed on acid-free paper

This Springer imprint is published by Springer Nature

The registered company is Springer International Publishing AG Switzerland

To my wife, Bai Rong.

– Wen Ming Liu

To my wife Quan, with love.

– Lingyu Wang

Preface¹

With rapid advancements in information technology, today's organizations routinely collect, store, analyze, and redistribute vast amounts of data about individuals, such as user account information and online activities. In addition, the next generation of smart systems (e.g., smart grids and smart medical devices) will enable organizations to collect personal data about every aspect of our daily life, from real-time power consumption to medical conditions.

Although collecting data may be essential for organizations to conduct their business, indiscriminate collection, retention, and dissemination of personal data represents a serious intrusion to the privacy of individuals. As a fundamental right of all individuals, privacy protection means organizations should only collect and retain sensitive personal information for purposes that have been agreed upon by the individuals and also keep collected information confidential and accessible only to authorized personnel.

Unfortunately, protecting personal information poses serious technical challenges in almost every stage of the data management life cycle, from data collection to data dissemination. A particularly insidious threat in this context is the *side-channel leak* in which an adversary makes inference of confidential data based on some seemingly innocent characteristics of the data, such as data packet sizes or knowledge about public algorithms used to generate the data. While side-channel attacks in specific domains, such as cryptosystems, are well studied, there exist little effort on generalizing side-channel attacks across different domains in order to understand their commonality.

This book studies side-channel leaks and corresponding countermeasures in several domains. First, we focus on privacy-preserving data publishing (PPDP) where side-channel leaks may be caused by adversaries' knowledge about the algorithms used to anonymize the data. For countermeasures, we first study a generic strategy independent of data utility measures and syntactic privacy properties, and then

¹Wen Ming Liu's work on this book was completed during his time as a Ph.D. student at Concordia University.

we propose a more efficient approach by decoupling privacy protection and utility optimization. Second, we examine Web applications where side-channel leaks may be caused by packet sizes and timing. For countermeasures, we first study a privacy-preserving traffic padding method inspired by the aforementioned PPDP solution, and then we further strengthen the approach against adversaries' external knowledge through random padding. Third, we look at smart metering where side-channel leaks may be caused by fine-grained meter readings. Finally, we discuss how those specific instances of side-channel leaks may be modeled using a generic model.

This book provides readers with not only detailed analysis of side-channel leaks and their solutions in each of the aforementioned domains but also a generic model that bridges the gaps between those different threats and solutions. The benefit of such knowledge is twofold. First, it provides readers with sufficient technical background to understand the threat of side-channel leaks in those domains and consequently exposes readers to many challenging and important issues that still remain attractive research topics today. Second, it can also lead readers to look beyond those three domains and apply the insights and ideas to derive novel solutions for dealing with side-channel leaks in other practical applications.

Montreal, QC, Canada

Lingyu Wang

Acknowledgments

This research was funded in part by the Natural Sciences and Engineering Research Council of Canada under Discovery Grant N01035.

Contents

| | | |
|----------|---|----|
| 1 | Introduction | 1 |
| 1.1 | Background | 1 |
| 1.2 | Overview | 2 |
| 1.3 | Summary of Contributions | 4 |
| | References | 6 |
| 2 | Related Work | 7 |
| 2.1 | Privacy Preservation | 7 |
| 2.2 | Side-Channel Attacks | 9 |
| 2.3 | Side-Channel Leaks in Data Publishing | 9 |
| 2.4 | Side-Channel Leaks in Web Applications | 11 |
| 2.5 | Side-Channel Leaks in Smart Metering | 11 |
| | References | 12 |
| 3 | Data Publishing: Trading Off Privacy with Utility Through the k-Jump Strategy | 17 |
| 3.1 | Overview | 17 |
| 3.2 | The Model | 21 |
| 3.2.1 | The Algorithms a_{naive} and a_{safe} | 22 |
| 3.3 | k -Jump Strategy | 24 |
| 3.3.1 | The Algorithm Family $a_{jump}(\mathbf{k})$ | 24 |
| 3.3.2 | Properties of $a_{jump}(\mathbf{k})$ | 26 |
| 3.4 | Data Utility Comparison | 28 |
| 3.4.1 | Data Utility of k -Jump Algorithms | 28 |
| 3.4.2 | Reusing Generalization Functions | 33 |
| 3.4.3 | The Relationships of a_{safe} and $a_{jump}(1)$ | 35 |
| 3.5 | Computational Complexity | 37 |
| 3.6 | Making Secret Choices of Algorithms | 39 |
| 3.6.1 | Secret-Choice Strategy | 39 |
| 3.6.2 | Subset Approach | 41 |
| 3.7 | Summary | 44 |
| | References | 44 |

| | |
|--|-----|
| 4 Data Publishing: A Two-Stage Approach to Improving Algorithm Efficiency | 45 |
| 4.1 Overview | 45 |
| 4.1.1 Motivating Example | 47 |
| 4.2 The Model | 49 |
| 4.2.1 The Basic Model | 49 |
| 4.2.2 <i>l</i> -Candidate and Self-Contained Property | 51 |
| 4.2.3 Main Results | 53 |
| 4.3 The Algorithms | 56 |
| 4.3.1 The RIA Algorithm (Random and Independent) | 57 |
| 4.3.2 The RDA Algorithm (Random and Dependent) | 59 |
| 4.3.3 The GDA Algorithm (Guided and Dependent) | 61 |
| 4.3.4 The Construction of SGSS | 63 |
| 4.4 Experiments | 63 |
| 4.4.1 Computation Overhead | 64 |
| 4.4.2 Data Utility | 65 |
| 4.5 Discussion | 67 |
| 4.6 Summary | 69 |
| References | 69 |
| 5 Web Applications: k-Indistinguishable Traffic Padding | 71 |
| 5.1 Overview | 71 |
| 5.2 The Model | 75 |
| 5.2.1 Basic Model | 75 |
| 5.2.2 Privacy and Cost Model | 76 |
| 5.3 PPTP Problem Formulation | 78 |
| 5.3.1 SVSD and SVM D | 79 |
| 5.3.2 MVMD | 80 |
| 5.4 The Algorithms | 82 |
| 5.5 Extension to <i>l</i> -Diversity | 84 |
| 5.5.1 The Model and Problem Formulation | 84 |
| 5.5.2 The Algorithms | 86 |
| 5.6 Evaluation | 89 |
| 5.6.1 Implementation and Experimental Settings | 89 |
| 5.6.2 Communication Overhead | 90 |
| 5.6.3 Computational Overhead | 92 |
| 5.6.4 Processing Overhead | 95 |
| 5.7 Summary | 95 |
| References | 96 |
| 6 Web Applications: Background-Knowledge Resistant Random Padding | 99 |
| 6.1 Overview | 99 |
| 6.1.1 Motivating Example | 101 |
| 6.2 The Model | 102 |
| 6.2.1 Traffic Padding | 103 |

- 6.2.2 Privacy Properties 103
- 6.2.3 Padding Method 105
- 6.2.4 Cost Metrics 106
- 6.3 The Algorithms 107
 - 6.3.1 The Random Ceiling Padding Scheme 107
 - 6.3.2 Instantiations of the Scheme 108
- 6.4 The Analysis 110
 - 6.4.1 Analysis of Privacy Preservation 110
 - 6.4.2 Analysis of Costs 113
 - 6.4.3 Analysis of Computational Complexity 116
- 6.5 Experiment 116
 - 6.5.1 Experimental Setting 116
 - 6.5.2 Uncertainty and Cost vs k 117
 - 6.5.3 Randomness Drawn from Bounded Uniform Distribution ... 117
 - 6.5.4 Randomness Drawn from Normal Distribution 118
- 6.6 Summary 121
- References 123
- 7 Smart Metering: Inferences of Appliance Status from Fine-Grained Readings** 125
 - 7.1 Overview 125
 - 7.2 Motivating Example 126
 - 7.3 The Model 128
 - 7.3.1 Adversary Model 129
 - 7.3.2 Privacy Property 129
 - 7.3.3 Cost Metrics 131
 - 7.4 Summary 132
 - References 132
- 8 The Big Picture: A Generic Model of Side-Channel Leaks** 133
 - 8.1 A General Framework for Side Channel Leaks 133
 - 8.1.1 Application Model 133
 - 8.1.2 Privacy Properties 135
 - 8.1.3 Cost Metrics 136
 - 8.1.4 Obfuscating Mechanisms 137
 - 8.2 Instantiations of the General Model 137
 - 8.2.1 Privacy-Preserving Data Publishing 138
 - 8.2.2 Privacy-Preserving Traffic Padding 139
 - 8.2.3 Privacy-Preserving Smart Metering 140
 - 8.2.4 Others 140
 - 8.3 Conclusion 141
 - 8.4 Future Work 142

Chapter 1

Introduction

1.1 Background

The privacy preserving issue has attracted significant attentions in various domains, including census data publishing, data mining, location-based services, mobile and wireless networks, social networks, Web applications, smart grids, and so on. A rich literature exists on this topic, with various privacy properties, utility measures, and privacy-preserving solutions developed. However, one of the most challenging threats to privacy, side-channel leaks, has received limited attention. In a side-channel leak, adversaries attempt to steal sensitive information not only from obvious sources, such as published data or the content of network packets, but also through other, less obvious (side) channels, such as their knowledge about anonymization algorithms or the packet sizes (to be discussed in more details in the coming chapters). Side channel leaks can usually further complicate privacy preservation tasks to a significant extent, as we will demonstrate in this book. Various side-channel attacks have been studied in different domains, such as:

- data publishing (e.g., adversarial knowledge about a generalization algorithm may allow adversaries to potentially infer more sensitive information from the disclosed data);
- Web-based Application (e.g., exact user inputs can potentially be inferred from the packet sizes even if the traffic between client and server sides is encrypted);
- smart metering (e.g., the fine-grained meter readings may be used to track the appliance’s usage patterns and consequently sensitive information about the household, such as daily activities or individuals’ habits);
- cloud computing (e.g., the sharing of physical infrastructure among tenants allows adversaries to extract sensitive information about other tenants’ co-resident VMs);