Advances in Information Security 68

Wen Ming Liu
Lingyu Wang

Preserving Privacy Against Side-Channel Leaks

From Data Publishing to Web Applications

Springer

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

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

Preface[1]

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

[1] Wen 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

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);

W.M. Liu, L. Wang, *Preserving Privacy Against Side-Channel Leaks*,
Advances in Information Security 68, DOI 10.1007/978-3-319-42644-0_1