

# SpringerBriefs on Cyber Security Systems and Networks


## Editor-in-Chief


Yang Xiang, Digital Research & Innovation Capability Platform, Swinburne University of Technology, Hawthorn, VIC, Australia

## Series Editors

Liqun Chen , Department of Computer Science, University of Surrey, Guildford, Surrey, UK

Kim-Kwang Raymond Choo , Department of Information Systems, University of Texas at San Antonio, San Antonio, TX, USA

Sherman S. M. Chow , Department of Information Engineering, The Chinese University of Hong Kong, Hong Kong, Hong Kong

Robert H. Deng , School of Information Systems, Singapore Management University, Singapore, Singapore

Dieter Gollmann, E-15, TU Hamburg-Harburg, Hamburg, Hamburg, Germany

Kuan-Ching Li, Department of Computer Science & Information Engineering, Providence University, Taichung, Taiwan

Javier Lopez, Computer Science Department, University of Malaga Computer Science Dept., Malaga, Spain

Kui Ren, University at Buffalo, Buffalo, NY, USA

Jianying Zhou , Infocomm Security Department, Institute for Infocomm Research, Singapore, Singapore

The series aims to develop and disseminate an understanding of innovations, paradigms, techniques, and technologies in the contexts of cyber security systems and networks related research and studies.

It publishes thorough and cohesive overviews of state-of-the-art topics in cyber security, as well as sophisticated techniques, original research presentations and in-depth case studies in cyber systems and networks. The series also provides a single point of coverage of advanced and timely emerging topics as well as a forum for core concepts that may not have reached a level of maturity to warrant a comprehensive textbook.

It addresses security, privacy, availability, and dependability issues for cyber systems and networks, and welcomes emerging technologies, such as artificial intelligence, cloud computing, cyber physical systems, and big data analytics related to cyber security research. The main focus is on the following research topics:

#### *Fundamentals and theories*

- Cryptography for cyber security
- Theories of cyber security
- Provable security

#### *Cyber Systems and Networks*

- Cyber systems security
- Network security
- Security services
- Social networks security and privacy
- Cyber attacks and defense
- Data-driven cyber security
- Trusted computing and systems

#### *Applications and others*

- Hardware and device security
- Cyber application security
- Human and social aspects of cyber security

More information about this series at <https://link.springer.com/bookseries/15797>

Jin Li · Ping Li · Zheli Liu · Xiaofeng Chen ·  
Tong Li

# Privacy-Preserving Machine Learning

Jin Li  
School of Computer Science and Cyber  
Engineering, Institute of Artificial  
Intelligence and Blockchain  
Guangzhou University  
Guangzhou, Guangdong, China

Zheli Liu  
College of Cyber Science and College  
of Computer Science  
Nankai University  
Tianjin, China

Tong Li  
College of Cyber Science and College  
of Computer Science  
Nankai University  
Tianjin, China

Ping Li  
School of Computer Science  
South China Normal University  
Guangzhou, Guangdong, China

Xiaofeng Chen  
State Key Laboratory of Integrated Service  
Network  
Xidian University  
Xi'an, China

ISSN 2522-5561                      ISSN 2522-557X (electronic)  
SpringerBriefs on Cyber Security Systems and Networks  
ISBN 978-981-16-9138-6              ISBN 978-981-16-9139-3 (eBook)  
<https://doi.org/10.1007/978-981-16-9139-3>

© The Author(s), under exclusive license to Springer Nature Singapore Pte Ltd. 2022

This work is subject to copyright. All rights are solely and exclusively licensed 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, expressed 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.

This Springer imprint is published by the registered company Springer Nature Singapore Pte Ltd.  
The registered company address is: 152 Beach Road, #21-01/04 Gateway East, Singapore 189721, Singapore

# Preface

As an implementation methodology of the artificial intelligence, machine learning techniques have reported impressive performance in a variety of application domains, such as risk assessment, medical predictions, and face recognition. Due to critical security concerns, how to protect data privacy in machine learning tasks has become an important and realistic issue spanning multiple disciplines. An ever-increasing number of researches have started proposing countermeasures to mitigate the threats of privacy leaks.

After motivating and discussing the meaning of privacy-preserving techniques, this book is devoted to provide a thorough overview of the evolution of privacy-preserving machine learning schemes over the last 10 years. We report these works according to different learning tasks.

In a learning task, a natural question is how the participants take the advantage of cooperative learning (Chap. 2) on the joint dataset of all participants' data while keeping their data private. Alternatively, besides the basic security requirements, the participants could suffer some bottlenecks on resources of computation, communication, and storage. Thus, they can outsource their computation workloads to cloud servers and enjoy the unlimited computation resources in a secure outsourced learning (Chap. 3) manner.

Massive data collection required for large-scale deep learning not only presents obvious privacy issues but also introduces the problems of efficiency and parallelization. The framework of distributed federated learning (Chap. 4) is necessary, by which the optimization algorithms used in deep learning can be parallelized and executed asynchronously. Moreover, to prevent learning results exposing private individual information in the dataset, the federated learning algorithm is supposed to achieve the differential privacy (Chap. 5) which is a strong standard for privacy guarantees for random algorithms on aggregate datasets.

Nowadays, machine learning classification is used for many data-driven applications. So, it is important to consider secure inference techniques (Chap. 6), in which the data and the classifier remain confidential when a user queries a classifier not owned by him/her. In Chap. 7, we turn to a concrete application, *i.e.*, privacy-preserving image processing.

This book is meant as a thorough introduction to the problems and techniques but is not intended to be an exhaustive survey. We can cover only a small portion of works of privacy-preserving machine learning.

Guangzhou, China

Guangzhou, China

Tianjin, China

Xi'an, China

Tianjin, China

Jin Li

Ping Li

Zheli Liu

Xiaofeng Chen

Tong Li

# Contents

<b>1</b>	<b>Introduction</b>	1
1.1	What Is Machine Learning?	1
1.2	Why Machine Learning Needs Privacy-Preserving Manner	4
1.3	Security Threats	7
1.4	Bibliographic Notes	9
	References	11
<b>2</b>	<b>Secure Cooperative Learning in Early Years</b>	15
2.1	An Overview of Neural Network	15
2.2	Back-Propagation Learning	17
2.3	Vertically Partitioned Training Dataset	20
2.3.1	Privacy-Preserving Two-Party Training	20
2.3.2	Secure Manner	21
2.3.3	Scheme Details	21
2.3.4	Analysis of Security and Accuracy Loss	24
2.4	Arbitrarily Partitioned Training Dataset	25
2.4.1	BGN Homomorphic Encryption	26
2.4.2	Overviews	26
2.4.3	Scheme Details	27
	References	30
<b>3</b>	<b>Outsourced Computation for Learning</b>	31
3.1	Outsourced Computation	31
3.2	Multi-key Privacy-Preserving Deep Learning	32
3.2.1	Deep Learning	32
3.2.2	Homomorphic Encryption with Double Decryption Mechanism	35
3.2.3	Basic Scheme	37
3.2.4	Advance Scheme	39
3.2.5	Security Analysis	43
	References	45

- 4 Secure Distributed Learning** ..... 47
  - 4.1 Distributed Privacy-Preserving Deep Learning ..... 47
    - 4.1.1 Distributed Selective SGD ..... 48
    - 4.1.2 Scheme Details ..... 49
  - 4.2 Secure Aggregation for Deep Learning ..... 51
    - 4.2.1 Secure Manner ..... 52
    - 4.2.2 Technical Intuition ..... 53
    - 4.2.3 Secure Protocol ..... 54
  - References ..... 56
- 5 Learning with Differential Privacy** ..... 57
  - 5.1 Differential Privacy ..... 57
    - 5.1.1 Definition ..... 57
    - 5.1.2 Privacy Mechanism ..... 58
  - 5.2 Deep Learning with Differential Privacy ..... 60
    - 5.2.1 Differentially Private SGD Algorithm ..... 60
    - 5.2.2 Privacy Account ..... 61
  - 5.3 Distributed Deep Learning with Differential Privacy ..... 62
    - 5.3.1 Private Algorithm ..... 62
    - 5.3.2 Estimating Sensitivity ..... 63
  - References ..... 64
- 6 Applications—Privacy-Preserving Image Processing** ..... 65
  - 6.1 Machine Learning Image Processing for Privacy Protection ..... 65
  - 6.2 Feature Extraction Methods of Machine Learning Image Processing ..... 66
  - 6.3 Main Models of Machine Learning Image Processing for Privacy Protection ..... 67
    - 6.3.1 Privacy-Preserving Face Recognition ..... 68
    - 6.3.2 Privacy-Preserving Object Recognition ..... 70
    - 6.3.3 Privacy-Preserving Classification ..... 72
  - Reference ..... 74
- 7 Threats in Open Environment** ..... 75
  - 7.1 Data Reconstruction Attack ..... 75
    - 7.1.1 Threat Model ..... 76
    - 7.1.2 Attack Method ..... 77
  - 7.2 Membership Inference Attack ..... 79
    - 7.2.1 Threat Model ..... 79
    - 7.2.2 Attack Method ..... 80
  - 7.3 Model Stealing Attack ..... 83
    - 7.3.1 Threat Model ..... 84
    - 7.3.2 Attack Method ..... 85
  - References ..... 86
- 8 Conclusion** ..... 87