

# DATA PRIVACY AND PRIVACY ENGINEERING

BY

AHMAD IZZUDDIN ZUHRI BIN ZAIDIN

A dissertation submitted in fulfillment of the requirement for  
the degree of Master of Protective Security Management.

Kulliyyah of Information Technology  
International Islamic University Malaysia

AUGUST 2022

## ABSTRACT

Due to the Covid-19 Pandemic outbreaks, many organizations have transformed their business processes from conventional-ways to digital platforms and approach for better and efficient communication and information systems. Subsequently, organizations that primarily use information systems when performing business processes will deal with numerous Personal Identifiable Information (PII) provided by clients. Moreover, today, most individuals use IT products and services-in which their PII are being collected and further processed. Due to that, PII must always be protected to avoid privacy risks and abused. However, concerns on data privacy among public is still low due to lack of understanding and knowledge on data privacy implication to ones' safety and security. Apart from that, software engineers or system owners faced several difficulties to determine the proper tools to be used in order to protect PII. This is related to the lack of awareness on Privacy Enhancing Technologies (PETs). This research attempts to close the gap of understanding on privacy data and introduce the discipline of Privacy Engineering. This research highlights relevant and important facts for public awareness and concern and provides a summary of PETs for software engineers and system owners. Method used to achieve the first objective is by conducting literature review to identify the relevant and reliable references about data privacy. Another method used is by studying and listing out available PETs' tools and approaches to facilitate software engineers in determining proper PETs to be applied. Collections and categorizations of description and summary of a data privacy elements for public understanding is the result produced by the first method. The second method produced a compilation and summary of PETs' available tools and techniques and an explanation on PETs adoption guide that can be applied by software engineers. This piece of research work may help public to understand on the importance of having a good protection of their PII and helps software engineers and system owners to be aware of proper tools to be used to enhance the protection of PII in their systems.

## ملخص البحث

بسبب تفشي جائحة كورونا **Covid-19** ، حولت العديد من المؤسسات عملياتها التجارية من الطرق التقليدية إلى منصات ونهج رقمية أفضل وأكثر كفاءة لأنظمة الاتصالات والمعلومات. في وقت لاحق، تتعامل المنظمات التي تستخدم أنظمة المعلومات بشكل أساسي عند تنفيذ العمليات التجارية مع العديد من معلومات التعريف الشخصية (**PII**) **Personal Identifiable Information** التي يقدمها العملاء. بالإضافة إلى ذلك، يستخدم معظم الأفراد اليوم منتجات وخدمات تكنولوجيا المعلومات التي يتم فيها جمع معلومات التعريف الشخصية الخاصة بهم ومعالجتها بشكل أكبر. ونتيجة لذلك، يجب دائما حماية معلومات التعريف الشخصية لتجنب المخاطر التي تهدد الخصوصية وإساءة الاستخدام. ومع ذلك، لا تزال المخاوف بشأن خصوصية البيانات بين الجمهور منخفضة بسبب نقص الوعي والمعرفة بشأن خصوصية البيانات التي تؤثر على سلامة وأمن الأشخاص. وبصرف النظر عن ذلك، واجه مهندسو البرمجيات أو مالكو الأنظمة العديد من الصعوبات لتحديد الأدوات المناسبة لاستخدامها لحماية معلومات التعريف الشخصية. ويرتبط ذلك بنقص الوعي بتقنيات تعزيز الخصوصية **Privacy Enhancing Technologies (PETs)**. يحاول هذا البحث سد الفجوة في فهم بيانات الخصوصية وتقديم تخصص هندسة الخصوصية. يسلط هذا البحث الضوء على الحقائق ذات الصلة والهامة للوعي العام والاهتمام ويقدم ملخصا لتقنيات تعزيز الخصوصية لمهندسي البرمجيات ومالكي الأنظمة. الطريقة المستخدمة

لتحقيق الهدف الأول هي إجراء مراجعة للدراسات السابقة لتحديد المراجع ذات الصلة والموثوقة حول خصوصية البيانات. وهناك طريقة أخرى مستخدمة تتمثل في دراسة وإدراج أدوات ونهج تقنيات تعزيز الخصوصية المتاحة لتسهيل مهندسي البرمجيات في تحديد **PETs** المناسبة التي سيتم تطبيقها. حيث تنتج الطريقة الأولى مجموعات وتصنيفات الوصف وملخص عناصر خصوصية البيانات لفهم الجمهور.

وأنتجت الطريقة الثانية تجميعاً وموجزاً للأدوات والتقنيات المتاحة لتقنيات تعزيز الخصوصية وشرحاً لدليل اعتماد **PETs** الذي يمكن أن يطبقه مهندسو البرمجيات. قد يساعد هذا العمل البحثي الجمهور على فهم أهمية وجود حماية جيدة لمعلومات التعريف الشخصية الخاصة بهم ويساعد مهندسي البرمجيات ومالكي الأنظمة على أن يكونوا على دراية بالأدوات المناسبة التي يجب استخدامها لتعزيز حماية معلومات التعريف الشخصية في أنظمتهم.



ASST. PROF. DR. ZAINAB SENAN ATTAR BASHI  
Assistant Professor  
Department of Computer Science  
Kulliyah of Information and Communication Technology  
International Islamic University Malaysia

## APPROVAL PAGE

I certify that I have supervised and read this study and that in my opinion, it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Protective Security Management

.....  
Normaziah Abd Aziz  
Supervisor

I certify that I have read this study and that in my opinion it conforms to acceptable standards of scholarly presentation and is fully adequate, in scope and quality, as a dissertation for the degree of Master of Protective Security Management

.....  
Hafizah Mansor  
Examiner

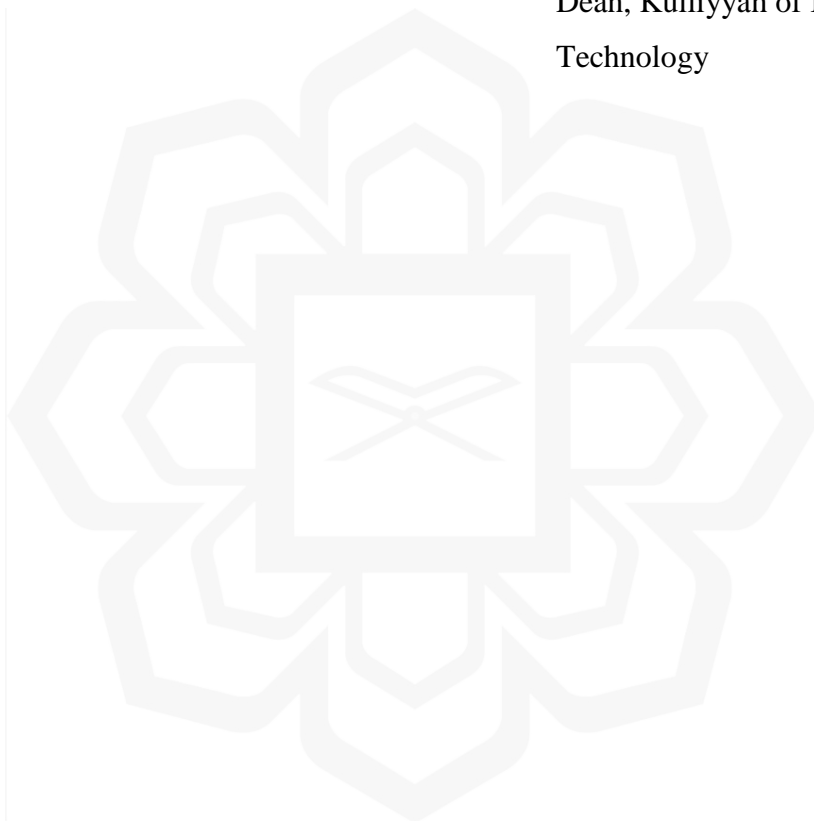
This dissertation was submitted to the Department of Centre For IT Advancement (CITA) and is accepted as a fulfilment of the requirement for the degree of Master of Protective Security Management

.....  
Madiah S. Abd. Aziz  
Head, Department of Kulliyah  
Information Technology

This dissertation was submitted to the Kulliyah of Information & Communication Technology and is accepted as a fulfillment of the requirement for the degree of Master of Protective Security Management

.....

Abdul Rahman Ahlan  
Dean, Kulliyah of Information  
Technology



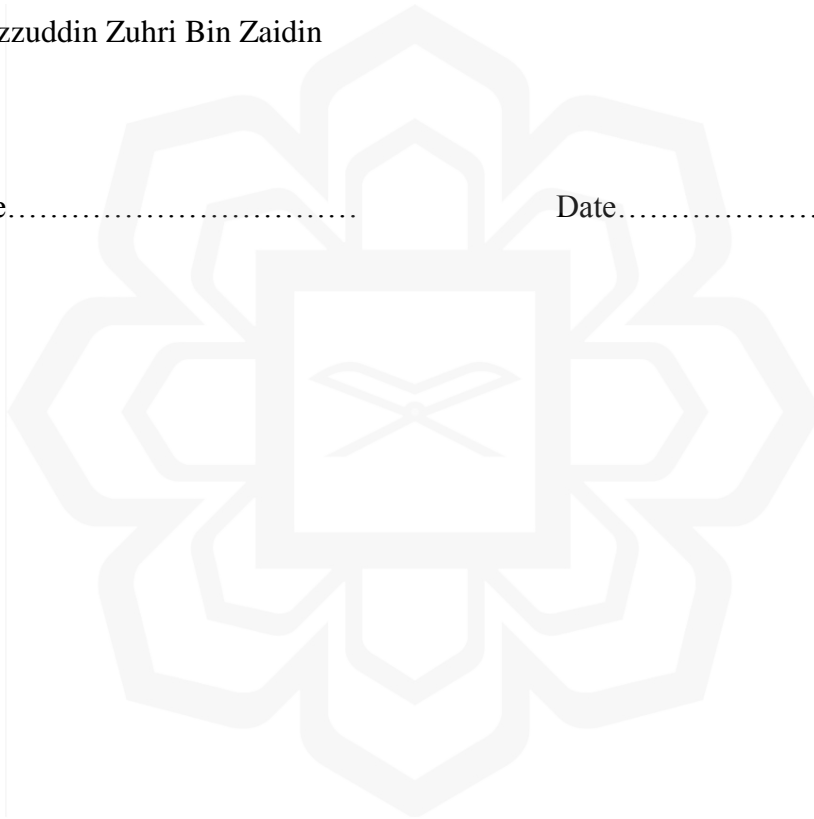
## DECLARATION

I hereby declare that this dissertation is the result of my own investigations, except where otherwise stated. I also declare that it has not been previously or concurrently submitted as a whole for any other degrees at IIUM or other institutions.

Ahmad Izzuddin Zuhri Bin Zaidin

Signature.....

Date.....



**INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA**

**DECLARATION OF COPYRIGHT AND AFFIRMATION OF  
FAIR USE OF UNPUBLISHED RESEARCH**

**TITLE OF THE THESIS/DISSERTATION**

I declare that the copyright holder of this dissertation are jointly owned by the student and IIUM.

Copyright © 2022 Ahmad Izzuddin Zuhri Bin Zaidin and International Islamic University Malaysia. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.
3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Ahmad Izzuddin Zuhri Bin Zaidin

.....

Signature

.....

Date



**INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA**

**DECLARATION OF COPYRIGHT AND AFFIRMATION OF  
FAIR USE OF UNPUBLISHED RESEARCH**

**TITLE OF THE THESIS/DISSERTATION**

I declare that the copyright holder of this dissertation is International Islamic University Malaysia.

Copyright © 2022 International Islamic University Malaysia. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below

1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.
3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Ahmad Izzuddin Zuhri Bin Zaidin

.....

Signature

.....

Date

**INTERNATIONAL ISLAMIC UNIVERSITY MALAYSIA**

**DECLARATION OF COPYRIGHT AND AFFIRMATION OF  
FAIR USE OF UNPUBLISHED RESEARCH**

**TITLE OF THE THESIS/DISSERTATION**

I declare that the copyright holder of this dissertation is Ahmad Izzuddin Zuhri Bin Zaidin.

Copyright © 2022 Ahmad Izzuddin Zuhri Bin Zaidin. All rights reserved.

No part of this unpublished research may be reproduced, stored in a retrieval system, or transmitted, in any form or by any means, electronic, mechanical, photocopying, recording or otherwise without prior written permission of the copyright holder except as provided below


1. Any material contained in or derived from this unpublished research may only be used by others in their writing with due acknowledgement.
2. IIUM or its library will have the right to make and transmit copies (print or electronic) for institutional and academic purpose.
3. The IIUM library will have the right to make, store in a retrieval system and supply copies of this unpublished research if requested by other universities and research libraries.

By signing this form, I acknowledged that I have read and understand the IIUM Intellectual Property Right and Commercialization policy.

Affirmed by Ahmad Izzuddin Zuhri Bin Zaidin

.....  
Signature

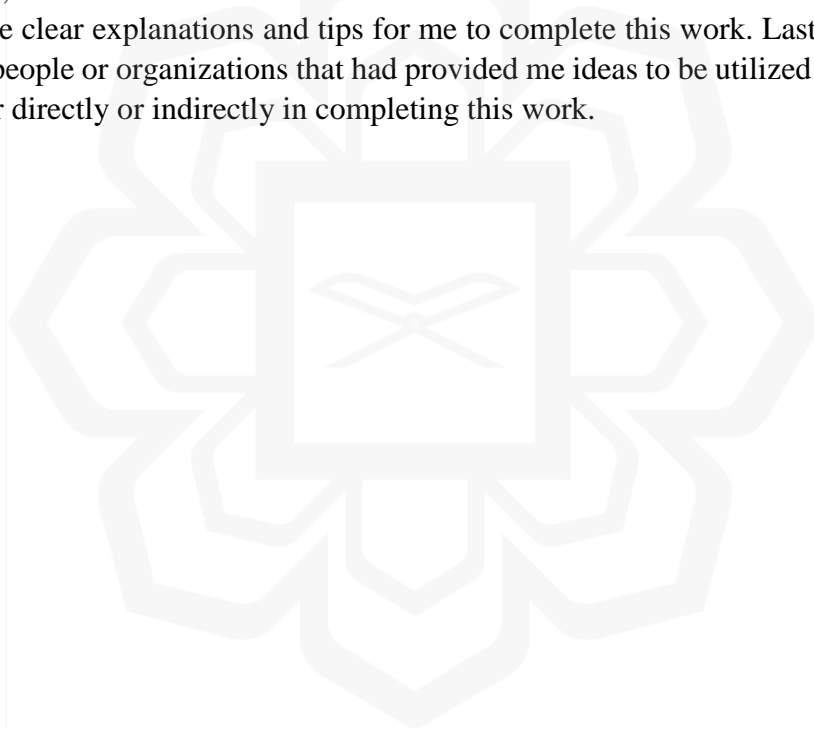
.....  
Date



*This Dissertation is dedicated to my parents for supporting me to become successful in  
life.*

## **ACKNOWLEDGEMENTS**

First of all, I would like to praise and thank Allah SWT for enabling and giving me some strength, motivation, and guidance to complete this work. Without his blessings, I could not successfully complete this work. Second, I would like to thank my supervisor, Assoc. Prof. Dr. Normaziah Bt. Abd Aziz for facilitating me to successfully complete this work through her motivation, guidance, and all sort of knowledge given by her. I would like to express my gratitude to my parents for their support for me to complete this work. Apart from that, I would like to thank the IIUM Centre for PostGraduate staff and CITA staff for giving me clear explanations and tips for me to complete this work. Lastly, special thanks to other people or organizations that had provided me ideas to be utilized hence had helped me either directly or indirectly in completing this work.



## TABLE OF CONTENTS

Abstract .....	ii
Abstract in Arabic .....	iii
Approval Page .....	vi
Declaration .....	viii
Copyright .....	ix
Dedication .....	xii
Acknowledgement .....	xiii
List of Tables .....	xix
List of Figures .....	xxi
List of Statues .....	xxv
List of Symbols .....	xxvi
List of Abbreviations .....	xxvii
<b>CHAPTER ONE: INTRODUCTION .....</b>	<b>1</b>
<b>CHAPTER TWO: LITERATURE REVIEW .....</b>	<b>11</b>
Definition of Privacy .....	11
Global Initiatives on Regulating Data Privacy List of Abbreviations .....	12
General Data Protection Regulation (GDPR) .....	13
California Consumer Privacy Act (CCPA).....	14
Personal Data Privacy Act 2010 Malaysia (PDPA).....	14
Privacy Act 1988 Australia .....	16
Real Case Incidents .....	18
Zoom Secretly Showed Personal Data From Linkedin Profiles .....	18

The Leakages Of 637,138 Of Albanian Citizens’ Personal Data .....	20
Google \$56 Million Fines For Privacy Breach .....	21
Facebook’S Facial Recognition Suit Violated Users’ Privacy .....	23
Summary of Real Case Incidents .....	24
Privacy By Design .....	25
Operationalizing Privacy By Design: A Guide To Implementing Strong Privacy Practices .....	26
NIST Privacy Framework .....	30
Deloitte Privacy Guidelines And Audit Strategies .....	35
ISO Privacy Framework .....	38
Summary .....	39
<b>CHAPTER THREE: PRIVACY ENGINEERING .....</b>	<b>40</b>
Privacy Engineering Features .....	41
Anonymization .....	42
Anonymization Techniques .....	44
Suppression .....	44
Generalization .....	45
Permutation or Randomization .....	46
Anatomization .....	47
Perturbation or Differential Privacy .....	51
k-anonymization .....	53
Pseudonymization .....	55
Comparisons Between Pseudonymization and Anonymization .....	57
Pseudonymization Techniques .....	58
Traditional Pseudonymization .....	58
Hashing without Key .....	60
Hashing with a Key .....	61
Encryption .....	63
Asymmetric Encryption .....	64

Tokenization .....	65
Undetectability .....	66
Undetectability Techniques .....	67
Steganography .....	67
Dummy Traffic .....	68
Spread Spectrum .....	68
Unobservability .....	69
Unobservability Techniques .....	70
Communication Anonymizer .....	70
Mix Network .....	70
Homomorphic Encryption .....	71
Trusted Execution Environment (TEE) .....	72
Unlinkability .....	73
Unlinkability Techniques .....	74
Federated Learning and Analytics .....	75
Multi-party Computing .....	76
Blind Signatures .....	78
Updatable Anonymous Credential System (UACS).....	78
Real-World Use Cases Utilized Emerging (PETs) .....	79
PETs Adoption Guidelines .....	83
Summary .....	86

**CHAPTER FOUR: PRIVACY ENHANCING TECHNOLOGIES (PETs) TOOLS AND APPROACHES ..... 87**

Privacy Enhancing Technologies Tools or Approaches .....	87
Very Good Security (VGS).....	87
How VGS Works .....	89
Libelle DataMasking Software .....	96
How Libelle DataMasking Works .....	96
ARX .....	103

How ARX Works .....	108
Brighter AI .....	112
How Brighter AI Works .....	113
Azure Confidential Computing (Virtual Machine).....	119
Amnesia .....	120
How AmnesiaWorks .....	121
Tensorflow Federated .....	126
LINDDUN .....	129
Private AI .....	135
How Private AI Works .....	136
Polymorphic Encryption and Pseudonymization (PEP) .....	137
How PEP Works .....	139
WEB-MPC Data Aggregation Approach.....	141
Privacy Preserving Loyalty points (PLPP) .....	143
How PLPP Works .....	144
Functionality of Hiding Database and Loyalty Points .....	144
Microsoft SEAL .....	146
g9 Anonymizer .....	148
How g9 Anonymizer Works .....	150
ProtonMail .....	158
How ProtonMail Works .....	158
End-To-End Encryption .....	160
Zero Access to User Data .....	161
Open Source Cryptography .....	161
Google Cloud DLP .....	162
Anonymization processes and How it Works .....	164
Pseudonymization processes and How it Works .....	164
Evaluation Table .....	169
Summary .....	173



<b>CHAPTER FIVE: PRIVACY REQUIREMENTS COMPARISONS BETWEEN COMPANIES .....</b>	<b>176</b>
IT Solutions Table .....	176
Social Media Companies .....	178
E-commerce Companies .....	180
Healthcare .....	183
Summary .....	185
<b>CHAPTER SIX: CONCLUSION AND RECOMMENDATIONS .....</b>	<b>187</b>
Overall Conclusions .....	187
Research Background Summary .....	189
Improvement and Future Research .....	190
<b>REFERENCES .....</b>	<b>192</b>

## LIST OF TABLES

Table 2.1	Identify - P	32
Table 2.2	Govern - P	33
Table 2.3	Communicate - P	34
Table 2.4	Process of assessing client's privacy requirements	37
Table 3.1	Example of anonymization process	46
Table 3.2	Original table	48
Table 3.3	Sensitive attribute table	48
Table 3.4	Quasi identifiers table	48
Table 3.5	{Cancer type, Treatment} table	49
Table 3.6	{{Gender, Age}, {Zipcode}} table	49
Table 3.7	Complete anatomization techniques	50
Table 3.8	Original table	53
Table 3.9	k-anonymization of k=2	54
Table 3.10	k-anonymization of k=4	54
Table 3.11	Comparisons between pseudonymization and anonymization	57-58
Table 3.12	Example of traditional pseudonymization	59
Table 3.13	Example of hashing without key	60
Table 3.14	Real world cases utilizing PETs	79-81
Table 4.1	ARX privacy models lists	105
Table 4.2	Transformation models lists	106
Table 4.3	DFD element over LINDDUN privacy threats	131

Table 4.4	Selected LINDDUN privacy threats	134
Table 4.5	Selection of proper privacy enhancing technologies and mitigation strategies	135
Table 4.6	Description on each view	152
Table 4.7	Original data	157
Table 4.8	Anonymized data	157
Table 4.9	List of alphabet or character set name, radix, and character list	167
Table 4.10	Evaluations of identified PETs' tools and approaches	169- 172
Table 5.1	Compilation of Microsoft, Google, and Apple privacy statements	176- 178
Table 5.2	Compilation of Facebook and twitter privacy statement	178- 180
Table 5.3	Compilation of Lazada, Shopee, and Amazon privacy statements	180- 183
Table 5.4	Compilation of Malaysian healthcare and UK National Health Service privacy statements	183- 185
Table 6.1	Research background summary	189

## LIST OF FIGURES

Figure 2.1	Roles and responsibilities of individuals	28
Figure 2.2	Relations between core part and profiles part	35
Figure 2.3	An overview on how Deloitte Privacy by Design framework works	37
Figure 3.1	Example of how differential privacy works by using privacy loss perimeter ( $\epsilon$ )	52
Figure 3.2	Operation of cryptographic hash functions	61
Figure 3.3	Example of hashing with a key	62
Figure 3.4	Example of symmetric encryption technique	63
Figure 3.5	Example of asymmetric encryption technique	65
Figure 3.6	Example of mix networks	71
Figure 3.7	An overview on how TEE operates	73
Figure 3.8	An overview on how federated analytic works	76
Figure 3.9	An overview on how MPC works	77
Figure 3.10	Flowchart of PETs adoption guide	83
Figure 4.1	An overview on the VGS operation	89
Figure 4.2	Dashboard (home)	90
Figure 4.3	Specifying server	91
Figure 4.4	Determining the path	91
Figure 4.5	Determining the path	92
Figure 4.6	Setting the inbound route	92
Figure 4.7	Original sensitive data	93

Figure 4.8	Response box depicts the pseudonymization of sensitive data	94
Figure 4.9	Pseudonymization of sensitive data	94
Figure 4.10	Setting of outbound route	95
Figure 4.11	Artificial version of data is re-identified	95
Figure 4.12	Main masking menu	97
Figure 4.13	Masking profiles	98
Figure 4.14	Masking procedure and types of modules available	99
Figure 4.15	Inspection and check phase	100
Figure 4.16	Preparation phase	101
Figure 4.17	Anonymization phase	102
Figure 4.18	Review phase	103
Figure 4.19	Generalization hierarchies of values attribute “age”	104
Figure 4.20	Overview of perspectives of the graphical interface of the ARX Data Anonymization Tool	107
Figure 4.21	Import the datasets	108
Figure 4.22	Set types of attributes part	109
Figure 4.23	Create a generalization hierarchy part	109
Figure 4.24	Generalization hierarchies of age	110
Figure 4.25	Generalization hierarchies of phone numbers	110
Figure 4.26	Determine the differential privacy	111
Figure 4.27	The results of the anonymization process	111
Figure 4.28	Dashboard (home)	114
Figure 4.29	User photo upload	114
Figure 4.30	User determines presets	115
Figure 4.31	Example of the process	115
Figure 4.32	Original photo	116
Figure 4.33	Result of the deep anonymization process	117
Figure 4.34	Original photo	118

Figure 4.35	Result of the deep precision blur process	118
Figure 4.36	Explanation of Intel SGX enclaves	120
Figure 4.37	Example of the importing data process	121
Figure 4.38	Determine the delimiter	122
Figure 4.39	Determine type of attribute	122
Figure 4.40	Generating input data	123
Figure 4.41	Select data attribute	123
Figure 4.42	Result of the generalization hierarchy	124
Figure 4.43	Determination of dataset algorithm	124
Figure 4.44	Example of tree diagram of solution graph	125
Figure 4.45	The final result	126
Figure 4.46	TensorFlow Federated package installer manual	127
Figure 4.47	TensorFlow Federated installer manual	128
Figure 4.48	LINDDUN privacy threats categories	129
Figure 4.49	An overview on how LINDDUN operates	130
Figure 4.50	Data flow diagram (DFD)	131
Figure 4.51	Example of a threat tree of DFD element	132
Figure 4.52	Mitigation strategies taxonomy	133
Figure 4.53	Users determined the features	137
Figure 4.54	Original data and synthetic data	137
Figure 4.55	Pseudonymized data printed on test tube of blood sample	139
Figure 4.56	Phases of PEP approach	140
Figure 4.57	Web-MPC used in the Boston Women's Workforce Council portal	142
Figure 4.58	Functionality of HD focuses on interface hd.read	145
Figure 4.59	Part of loyalty points code	146
Figure 4.60	Differentiation between Microsoft SEAL cloud storage and traditional cloud storage	148
Figure 4.61	An overview on how g9 Anonymizer operates	150

Figure 4.62	Hotel Example Project	151
Figure 4.63	Example of anonymizer editor	152
Figure 4.64	Folder of the Company Table	153
Figure 4.65	Name column	154
Figure 4.66	Masking rule selection and the modification of Name column's property	155
Figure 4.67	Steps to execute console view	156
Figure 4.68	Console view	156
Figure 4.69	Home dashboard	159
Figure 4.70	Compose an email part	159
Figure 4.71	ProtonMail encryption method	160
Figure 4.72	Example of email received by others	162
Figure 4.73	Google Cloud home page	163
Figure 4.74	Deterministic encryption format	165
Figure 4.75	Pseudonymized sensitive data using deterministic encryption	165
Figure 4.76	Format preserving encryption format	166
Figure 4.77	Cryptographic hashing	168

## LIST OF STATUTES

California Consumer Privacy Act 2018 (California State)  
General Data Protection Regulations 2018 (European Union)  
Information Privacy Act 2014 (Australia)  
Personal Data Protection Act 2010 (Malaysia)  
The Privacy Act 1988 (Australia)

