MITIGATING PRIVACY RISKS OF SMARTPHONES IN MOBILE COMPUTING

LIU RUI

M.Phil

The Hong Kong Polytechnic University

2016

THE HONG KONG POLYTECHNIC UNIVERSITY
DEPARTMENT OF COMPUTING

MITIGATING PRIVACY RISKS OF SMARTPHONES IN MOBILE COMPUTING

LIU Rui

A thesis submitted in partial fulfillment of the requirements for
the degree of Master of Philosophy

June 2015

CERTIFICATE OF ORIGINALITY

I hereby declare that this thesis is my own work and that, to the best of my knowledge and belief, it reproduces no material previously published or written, nor material that has been accepted for the award of any other degree or diploma, except where due acknowledgement has been made in the text.

_____ (Signed)

LIU Rui (Name of Student)

Abstract

The ubiquitous and ever-more-capable smartphones bring forth unprecedented performance in mobile computing. The pursuit of high quality mobile applications and services may however compromise users' privacy, which is a pivotal issue in mobile computing. Much attention goes into how to mitigate users' privacy risks in smartphones. Compared with traditional privacy-preserving methods, new challenges have raised in smartphone privacy. On one hand, users have to provide their information for better functionality and service in the smartphone. On the other hand, they are reluctant to reveal some sensitive or personal data. In this thesis, we investigate smartphone privacy to address these new challenges. We survey the state of the art on the smartphone privacy, focusing on the current issues, proposed methods and existing systems. We discuss the characteristics of smartphone privacy in mobile computing and then review a number of related works and on-going research in detecting and mitigating privacy risks in smartphones. According to our findings, we consider two important cases of smartphone privacy disclosure in mobile computing: privacy leakage during mobile participatory sensing and privacy disclosure of mobile applications.

First, we study smartphone privacy during mobile participatory sensing with a focus on privacy measurement. With the development of mobile devices, a novel sensing paradigm emerges, namely, participatory sensing, which engages users with mobile devices to collect and interpret sensory information from the environment. The users participate in multifarious sensing tasks and share their information. It is not uncommon that their privacy is likely disclosed when the information is shared. Current works focus on privacy protecting and preserving and propose algorithms and mechanisms to prevent users' information from being disclosed. However, users are reluctant to hold their data perpetually since it is routine for them to share information in a participatory sensing systems. Users, therefore, need to know how much the privacy risk they have and which data can be shared. Unfortunately, it is arduous for users to apprehend their privacy risk in mobile participatory

sensing systems, and make a proper decision on data sharing accordingly. To address this issue, we propose a privacy measurement method, PriMe, which quantifies the privacy in participatory sensing systems from the perspective of individual sensitivity. Participants are recruited to conduct the experiments for evaluation. The experiment results show that PriMe can provide accurate results to the participants.

Second, we study smartphone privacy in mobile applications, focusing on mitigating users' privacy risks. Privacy is a crucial issue of mobile apps because there is a plethora of personal and sensitive information in smartphones. Various mechanisms and tools have been proposed to detect and mitigate privacy leaks. However, they rarely consider users' preferences and expectations. Users hold various expectations towards different mobile apps. For example, users may allow a social network app to access their photos rather than a game app because it is beyond users' expectation that an entertainment app attempt to get the personal photos. Therefore, it is vital to understand users' privacy expectations of various mobile apps and help them to mitigate privacy risks in the smartphone accordingly. To achieve this objective, we propose and implement PriWe, a system based on crowdsourcing driven by users who share privacy permission settings of their apps in smartphones. PriWe leverages the crowdsourced permission settings to understand users' privacy expectation and provides app specific recommendations to mitigate information leakage. We deployed PriWe in the real world for evaluation. Feedbacks are collected from the real world users and participants on Amazon Mechanical Turk. The results show that PriWe can make proper recommendations which meet participants' privacy expectation and are accepted by the users.

Publications

1. **Rui Liu**, Jiannong Cao, Sebastian VanSyckel, "*PriMe: Human-Centric Privacy Measurement based on Users' Preferences towards Data Sharing in Mobile Participatory Sensing Systems*", 14th IEEE International Conference on Pervasive Computing and Communications (PerCom 2016), March 14-18, 2016. Sydney, Australia.

2. **Rui Liu**, Jiannong Cao, Lei Yang, Kehuan Zhang, "*PriWe: Recommendation for Privacy Settings of Mobile Apps based on Crowdsourced Users' Expectations*", 4th IEEE International Conference on Mobile Services (IEEE MS 2015), June 27 - July 2, 2015, New York, USA.

3. **Rui Liu**, Jiannong Cao, Lei Yang, "*Smartphone Privacy in Mobile Computing: Issues, Methods and Systems*", DBSJ Journal, Vol.13, No.1, pp.1-13, March 2015.

4. Ruiyun Yu, **Rui Liu**, Xingwei Wang, Jiannong Cao, "*Improving data quality with an accumulated reputation model in participatory sensing systems*", Sensors 2014, vol. 14, iss. 3, pp. 5573-5594.

5. Shan Jiang, Junbin Liang, Jiannong Cao, **Rui Liu**, "*An Ensemble-Level Programming Model with Real-Time Support for Multi-Robot Systems*", IEEE International Conference on Pervasive Computing and Communications Demonstration Track (PerCom 2016 Demo), March 14-18, 2016. Sydney, Australia.

6. **Rui Liu**, Jiannong Cao, Kehuan Zhang, Wenyu Gao, Lei Yang, "*Understanding Mobile Users' Privacy Expectations: A Recommendation-based Method using Crowdsourcing*", under major revision, IEEE Transactions on Services Computing (TSC).

7. **Rui Liu**, Jiannong Cao, Weiping Zhu, Sebastian VanSyckel, Christian Becker, Junbin Liang, "*UIO-based Testbed Augmentation for Simulating a Large-Scale Intelligent*

Transportation System", under major revision, IEEE Transactions on Industrial Informatics (TII).

8. **Rui Liu**, Jiannong Cao, Kehuan Zhang, Wenyu Gao, Lei Yang, Junbin Liang, " *When Privacy Meets Usability: Unobtrusive Privacy Permission Recommendation System for Mobile Apps based on Crowdsourcing*", under review, IEEE Transactions on Services Computing (TSC).

9. Ruiyun Yu, Jiannong Cao, **Rui Liu**, Xingwei Wang, Wenyu Gao, "*RIM: Reputation-based Incentive Mechanism towards Quality-Oriented Location-Centric Participatory Sensing*", under review, IEEE Transactions on Mobile Computing (TMC).

Acknowledgements

To everyone in my MPhil journey-this is for you.

Prof. Jiannong Cao, only with your advice and support can I pursue this ambitious thesis and degree. Your cheerful personality and rigorous scholarship have changed my world. You always trained me to be a good researcher, from method to attitude. Your systematic guidance and valuable suggestions open my door to the research. Your broad knowledge, keen insight, and enormous enthusiasm on the research inspire me a lot and encourage me to keep going in my study. I consider myself very fortunate to have the opportunity to learn from you. Finally, I know you are always there to support me no matter what difficulties may come my way.

Prof. Christian Becker, Dr. Xiapu Luo, Dr. Kehuan Zhang, thanks for your brotherly support, research idea, and life advice. I know you are very busy but you always discuss with me from time to time, and give me many constructive comments in my research and great help in the paper writing. Your suggestions about my future really guide me to find my real inner self. I hope and believe you can reach high achievements in your research career.

Dr. Junbin Liang, thanks for our precious friendship; your are a good coordinator in our research group. Dr. Weiping Zhu, your help makes me adapt quickly to the group. Dr. Yin Yuan, it is my pleasure to meet you in my life. Dr. Sebastian Vansyckel, thanks for your kind help in my MPhil journey. Mr. Florian Klingler, you are a very cheerful man, hope you have a bright future.

My other academic siblings in the IMCL group, I treasure our joyous moments and collaboration: Dr. Xuefeng Liu, Dr. Peng Guo, Dr. Lei Yang, Dr. Tao Li, Dr. Zongjian He, Mr. Yang Liu, Mr. Guanqing Liang, Ms. Wanyun Lin, Mr. Junhao Zheng, Mr. Yaguang HuangFu and many other members I can not enumerate. Thank a lot for your help in these years. We learn from each other, share our joyfulness and sadness, and have a unforgettable memory together. I wish all of you a brilliant future.

I also want to thank my roommate Mr. Yu Xie, I am lucky to meet you in Hong Kong. Mr. Peng Liu, my another roommate, you make me feel at home when I escape from research.

Ms. Moyao Liu, Mr. Tao Li, Mr. Xiao Liu, Mr. Wengen Li, Mr. Zhaoxuan Ding, my dear friends, your every presence brightens my days.

Mom, dad, your endless love and care through out my life. No matter how far away, you are always my biggest and best supporters. I love you and miss you, though I never say it.

Table of Contents

List of Tables

List of Figures

List of Abbreviations

API: Application Program Interface

APK: Android Application Package

ART: Android Runtime

CFG: Control Flow Graph

DFA: Data Flow Analysis

GUI: Graphical User Interface

ICP: Inter-Process Communication

OS: Operating System

PM2.5: Particulate Matter smaller than 2.5 micrometers

UID: User ID

Chapter 1

Introduction

This research aims to investigate the issues of smartphone privacy in mobile computing and design novel mechanisms to mitigate the privacy risks of smartphones. In this chapter, we first describe the background knowledge of smartphone in mobile computing in Section 1.1. Then we introduce the smartphone privacy in Section 1.2. After that, we explain the motivation of our work in Section 1.3. In Section 1.4, we summarize the main contributions of this thesis. Finally, we outline the organization of this thesis in Section 1.5.

1.1 Smartphone in Mobile Computing

Mobile computing is a human computer interaction technology by which a computer is expected to be a mobile device that enables access to resources at any time, from any location [FZ94]. The typical mobile computing architecture for smartphone is shown in Fig. 1.1. There are several layers in the architecture. The smartphone application is the part which interacts with users and provides services directly. These applications are generated based on plentiful APIs and resources, which are provided by different mobile operating systems. The mobile systems are supported by increasing storage resources, powerful computing capacity, high-quality networks and sophisticated embedded sensors in hardware layer. The communication is another important function of smartphone in mobile computing. Under various kinds of network, users can access network, share data and receive information.

There are some innate flaws in this architecture, which may lead users' privacy leakages.

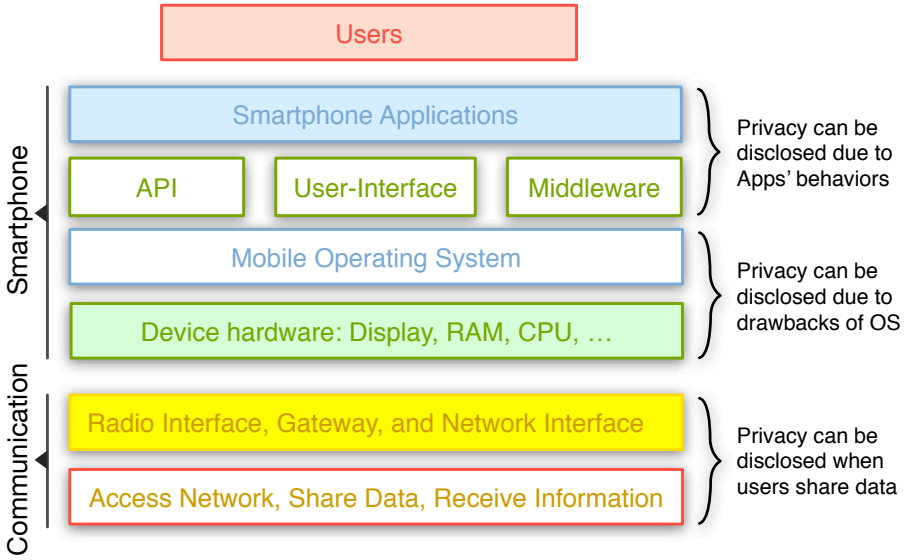


Fig. 1.1: The typical mobile computing architecture for smartphone

Firstly, the privacy disclosure can happen in smartphone application and mobile operating system. The operating system built in smartphones provides a plethora of applications aimed at making our life convenient [and14]. However, many smartphone applications, especially malicious ones in the mobile operating system can disclose privacy of smartphone users. A number of research have devoted themselves to these two layers, mobile operating system and mobile application, to address the privacy issues. Secondly, a substantial amount of personal information is exposed during the communication. For example, mobile sensing is a novel sensing paradigm which utilizes embedded mobile sensors to collect and share data. It is becoming mature and involved in our life. Unfortunately, it may disclose users' information because the people have no awareness what can be inferred from the sensory data and share their information to others. Thus, mobile sensing also has revealed the public privacy concerns.

1.2 Smartphone Privacy

We discuss some existing privacy definitions and describe several important characteristics of smartphone privacy in this section.

Privacy is by no means a fad of modern society. In 1890, Two U.S. lawyers proposed a prevalent definition, *privacy can be regarded as private life, habits, act, relations and the right to be alone* [WB90]. With the proliferation of information technology, Wesin proposed that *privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others*, and it came to be known as information privacy [Wes68]. These two acknowledged definitions both emphasized that privacy to users should be an ability to express themselves selectively. Moreover, as proposed by Bellotti and Sellen [BS93], privacy definition is not static and monolithic but should have different aspects due to new technology, patterns of use and social norms development.

In order to cater to the specific characteristics of smartphone, we express smartphone privacy from two perspectives, human-centric and technology-centric.

1.2.1 Human-centric Privacy

Based on previous works, a number of definitions consider people's concern as one of the most important factors. When people use smartphone, they have to provide their information for better services; meanwhile, they are reluctant to exposure sensitive data due to privacy concerns [SHC+09]. Therefore, human-centric smartphone privacy in mobile computing focuses on the balance between privacy and services [GB09], which mainly includes expectation, awareness and authorization.

- **Expectation** means people's expectation about how their information can be used by their smartphones. Namely, the data usage in smartphone should meet the users' expectation. For instance, users' expectation of a game application is entertainment rather than accessing many other kinds of information. Some game applications require a plethora of data, including users' accounts, approximate location information, personal photos and device ID, which may be not necessary for functioning [FHE+12]. The information abuse makes smartphones' behaviors ignore users' expectations. A number of works have been carried out to investigate users' expectations about their

information in the smartphones [SH14, LAH+12, BGS11].

- **Awareness** explains the degree of agreement between users' awareness and actual behaviors of smartphones. It is a curial issue since many free applications in mobile operating systems collected users' data without their awareness and intervention [FLL+13]. A lot of research works concentrate on this issue [ZXGC14, MS14] since users' awareness is significant in human-centric smartphone privacy.

- **Authorization** expresses that all decisions about usage of users' data in the smartphone, including removing, collecting, analyzing, publishing, should be made by themselves. Some researchers tried to built tools, systems and frameworks to protect users' privacy in the light of their own decisions [KCS13, BHS13].

Note that the ultimate goal of the research focused on human-centric smartphone privacy is to find balance between services and privacy. Namely, protecting users' privacy in smartphone is based on their attitudes, concerns, and preferences. It is the nature of human-centric privacy,

1.2.2 Technology-centric Privacy

Technology-centric smartphone privacy concentrates on protecting or preserving privacy through techniques according to different contexts for various goals, such as designing algorithms, developing mechanisms and building systems to prevent sensitive information from stealing and attacking. Technology-centric smartphone privacy mainly includes sensitivity and anonymity.

- **Sensitivity** refers to sensitive information stored in the smartphone. These information should be protected as much as possible. Users' privacy can be disclosed after obtaining or inferring some sensitive data [MPR08, LYL+10]. More particularly, the sensitive data may include time, location, acoustic and visual data, acceleration, environment context and biometric information.

- **Anonymity** usually considers the probability of users' data can be hidden after releasing [CKK+08]. Because completely protecting information sounds like an impossible mission, the advent of anonymity technology has aroused wide concerns. Some works like k-anonymity [Swe02], l-diversity [MKGV07] and t-closeness [LLV07] are proposed for guaranteeing the appropriate usage of data. Likewise, some research works try to achieve anonymity of smartphone data, currently most of them focus on location data in the smartphone [GL08, LHX13].

By comparison with human-centric smartphone privacy, the research on technology-centric privacy make more effort on how to protect information using technology without considering users' concerns and preferences. Furthermore, it should be noted that the technology-centric privacy is different from security in the smartphone. Security is usually referred to the confidentiality, availability, and integrity of data. Namely, the objective of security is to ensure the data is accurate, reliable and only accessed by authorized individuals or organizations. Technology-centric privacy however mainly focuses on achieving appropriate usage of data in smartphones through methods, tools and systems.

1.3 Motivations of Our Work

To preserve smartphone privacy in mobile computing, we need to consider the requirements discussed in Section 1.1 and Section 1.2. Although some related issues are well addressed, there are still many problems lacking sufficient investigation. In this section, we identify the problems which deserve further investigations.

The operating system is built in smartphones and further provides a plethora of applications aimed at making our life convenient [and14]. However, there may be some innate flaws in mobile operating systems, which can lead users' privacy leakage. As we know, iOS and Android are two of the most popular mobile operating systems, but many reports, news, investigations reveal that some drawbacks exist in the systems, which can be used by malicious users [OTK+12, FEW12, AH13].

Many smartphone applications, especially malicious ones, also bring the privacy risks.

Most applications are released as an encapsulated package or an executable program. From viewpoint of technology-centric privacy, it is arduous to learn what information can be disclosed. Therefore, an important issue is to detect and analyze the mobile applications so that we can know the potential privacy risks of them. In the Android system, each application has a permission list to claim what kind of data or function the app can use. The application will hold the permission when the users approve it. Actually, they have to agree it since it is a necessary step for installing the application. Thus, a lot of applications try to get as many permissions as they can, in this case the applications will lead permission abuse. For example, users' information can be stole or destroyed if some malicious applications can access personal data stored in the smartphones. It is also dangerous even normal application can get some unrelated data for application running. In iOS platform, we have privacy settings instead of the permission mechanism. However, the privacy settings also make users to face the similar problem. Thus, another issue is to address data abuse of smartphone apps. From human-centric privacy angle, individual concern to privacy differs from person to person, background to background. It is improper to assume that everyone has the same privacy preferences. Understanding users' privacy is becoming a problematic issue in smartphone privacy. The ultimate goal of smartphone privacy from application perspective is to balance the apps' functionality and users' privacy. The final issue is the tradeoff between utility and privacy.

The mobile communication is another vital function of the smartphones in mobile computing. Some works have paid more attention to privacy leakage when users share their data from connectivity perspective [Wic13, KKHK12]. In this field, mobile sensing is becoming popular due to the proliferation of sensor-equipped smartphones. It is a novel sensing paradigm which utilizes embedded mobile sensors to collect and share data. It is very different from some traditional privacy preserving scenarios. In mobile sensing, on one hand, it is not uncommon for users to share their information with others in mobile sensing tasks. On the other hand, they are reluctant to exposure sensitive data due to privacy concern. However, it is difficult for users to choose proper information for sharing

in appropriate participatory sensing tasks because they are unaware of what can be inferred from the sensory data. Hence, one issue is that how to share sensory data and preserve personal privacy as well.

In this thesis, we will analyze aforementioned problems in detail and propose corresponding solutions for them.

1.4 Contributions of the Thesis

The contributions of this thesis mainly lie in designing novel mechanisms for mitigating smartphone privacy in mobile computing. Our contributions include three parts:

First, we study privacy issues of smartphone users in terms of three aspects: mobile operating system, mobile application and mobile sensing, as shown in Fig. 1.1. Our main contributions in this part are: 1) investigating the characteristics of smartphone privacy; 2) pointing out the privacy issues according to the three aspects; 3) providing a taxonomy of existing methods and systems which can detect, analyze and mitigate privacy issues of smartphone users. Based on our survey, we identify two privacy issues as our further work, privacy measurement in mobile participatory sensing and mitigating privacy risks of smartphone applications.

Second, we focus on privacy measurement in mobile participatory sensing. Privacy in mobile participatory sensing is a fluid and malleable concept rather than a monolithic definition. Users show ambivalence with regard to the privacy. The users need to find the balance between the privacy and services. To address this non-trivial issue, we propose a privacy measurement method, PriMe, which quantifies the privacy in mobile participatory sensing systems. To the best of our knowledge, it is the first privacy measurement method in the participatory sensing systems. PriMe quantifies the privacy from the perspective of individual attitude, which is represented by two intuitive properties: the inherent sensitivity of each data item; and the individual sensitivity to each data item. Experimental results illuminate that PriMe provides accurate results to the participants.

Third, we focus on mitigating privacy risks of smartphone applications. The abuse of

permission is common in many Android applications. For instance, some game applications require a plethora of data, including users accounts, approximate location information, personal photos and device ID, which may be not necessary for functioning. The users' expectation of a game app is entertainment rather than accessing so many kinds of information. Obviously, the abuse of permission make applications behaviors beyond users' expectations, thus the risk of sensitive data disclosure is increased remarkably. To address this issue, it is significant to understand users expectation of privacy and mitigate abuse of data access permissions of Android applications. To achieve this objective, we propose PriWe, a system aims to understanding users' expectation of privacy and help them to make proper decisions based on their expectation. PriWe is first to learn the similarities among users in terms of privacy preferences and privacy expectations on apps, and then to recommend appropriate permission settings to users based on such similarities. The rationale behind our method is that: users who share similar preferences on certain private data and/or privacy expectations on apps are more likely to make similar decisions in related privacy items.

1.5 Organization of the Thesis

The structure of this thesis is organized as follows, Chapter 1 is the introduction to this thesis. Chapter 2 reviews related works in the literature. The main body of this thesis is divided into two parts in Chapter 3 to Chapter 4. The details are presented as follows.

In the first part, we mainly discuss our work about privacy measurement in participatory sensing. In Chapter 3, we presented PriMe, a privacy measurement method in participatory sensing systems. Based on the proposed properties of privacy in participatory sensing, we measure the privacy according to individual attitude, which is represented by the inherent sensitivity of each data item and the individual sensitivity to each data item. Experimental results illuminate that PriMe provides accurate results to participants when they are selecting sensing tasks in participatory sensing systems.

In the second part, we mainly discuss our work in mitigating privacy risk of smartphone

applications. In Chapter 4, we propose and implement PriWe, a system based on crowdsourcing driven by users who contribute privacy permission settings of their apps in smartphones. PriWe leverages the crowdsourced permission settings to understand users' privacy expectation and provides app specific recommendations to mitigate information leakage. We deployed PriWe in the real world for evaluation. According to the feedbacks of 78 users from the real world and 382 participants from Amazon Mechanical Turk, PriWe can make proper recommendations which can meet participants' privacy expectation and are mostly accepted by users, thereby help them to mitigate privacy disclosure in smartphones.

Finally, we conclude the thesis and discuss the directions of future works in Chapter 5.

Chapter 2

Literature Review

In this chapter, we review existing works about smartphone privacy in mobile computing. As we have discussed, we focus on three aspects, mobile operating system, mobile application, and mobile participatory sensing in this thesis. We first review the existing works about protecting privacy in mobile operating system in Section 2.1. Then we review the existing works about mitigating privacy risks of mobile application in Section 2.2. Finally, we review the existing works about preserving privacy in mobile participatory sensing in Section 2.3.

2.1 Existing Works about Protecting Privacy in Mobile Operating System

According to a recent report, Android has over 84% and iOS has 11.7% of the global market share in the third quarter of 2014 [smm]. The copious applications are produced based on the mobile systems and used by people everyday. These two popular systems both claim that they take users' privacy very seriously and retain the users' information just for better services [app14]. The users of Android and iOS however still faced the scads of privacy issues and risks. We therefore analyze these two mobile systems since they hold the biggest application market globally.

iOS platform

We discuss Apple iOS in this section, concentrating on its privacy features. For guaranteeing the privacy, some mechanisms such as code-signing, encryption and sandboxing are developed in iOS. In particularly, code-signing mechanism only allows code which is verified by Apple to run in the mobile devices. Encryption prevents the code from reverse-engineering and ensures the applications only can be launched by the purchasers. Sandboxing is designed for preserving users' privacy, preventing an application from accessing users' information in the smartphone. Moreover, iOS offers a gamut of APIs with developers to allow applications to communicate with each other using parameters.

However, many applications in iOS are designed to access shared information and resources, including sensitive data like location, photos, emails, contacts. What's worse, considering the tenet of iOS is to provide elegant and intuitive interface, the scads of interactions and technical details are hidden. For example, there is no alert or notification to users about their privacy when they are installing the applications. Recognizing the need for protecting user privacy, iOS introduces popup notifications. Users can set the permission for the data when an application want to access some personal information. However, according to some reports and survey, most people think the notification is intricate and few of them will read it [LBS+13].

Normally, iOS applications are distributed and reviewed via the App Store held by Apple. The review process currently includes static analysis and runtime analysis. They can make sure only authorized APIs are used and the applications would not obtain information by evading sandbox mechanism. However, it is very hard to scrutinize each application due to scads of iOS applications submitted. In fact, a number of malicious applications have passed the review process. Besides getting applications from App Store, another way is jailbreaking, which is certainly not supported by Apple. It is prevalent for normal users, even some research works about protecting user privacy are also based on jailbreaking [AH13]. In this case, the code-signing will be removed and applications can be installed from other sources, such as Cydia Store. We do not discuss the jailbreaking since it is out

of this thesis but it is obvious that jailbreaking iOS devices may cause more privacy risks.

The iOS is considered as a closed source system, there is few work that can improve privacy preserving of iOS except Apple itself. From iOS 4-8, privacy preserving technology and policy have been ameliorated, but the users' data are still in danger.

Android platform

Android is a Linux kernel based mobile operating system. It is designed for various mobile devices, especially smartphones. Its applications are written in Java and compiled into a custom byte code, which is known as ART and its predecessor Dalvik [MS12]. We briefly discuss Android mobile operating system and its ecosystem from a privacy perspective.

Protecting privacy in Android is based on sandbox, cryptography, secure IPC. Android applications run in a sandbox environment called Android Application Sandbox, which isolates a particular application's data and code execution from other applications, so that other processes on the system cannot access it. Cryptography and secure IPC actually include some implementations of common security functionality. More specifically, a wide array of algorithms using cryptography and an encrypted filesystem have been implemented to protect data. Meanwhile, Android provides plentiful APIs for developers to access local data in the smartphone and shared information by others [Mul10]. Although these APIs help applications to produce better services, most of accessed data are sensitive to users, including location, contacts, photos and so forth [COWC+13, KKHK12].

To address this issue, Android introduces permission mechanism in Android Play Store and puts notification in the smartphone. Permission mechanism allows applications to explicitly share resources and obtain additional capabilities not provided by the basic sandbox due to their needs [KBSW13]. All the applications in Play Store will show their data permissions before installing [HMN+13]. The data access notification will pop up when the application requires information. Unfortunately, few people will read the list and figure out why they need to hold such permissions, and most users just touch the accept button [KCC+12]. What is worse, mobile services cannot be provided sometimes due to the scarcity of required data if people reject those permissions [BJL+13]. Thus, compromise is the only

choice for users.

Furthermore, Google introduces Bouncer to automatically scrutinize applications to prevent malicious applications. However, a plethora of applications can circumvent the Bouncer and appear on the store. Beside Play Store, Android allows APK to distribute and install application software. This feature extremely expands application ecosystem and may lead more information disclosure [GZWJ12, VVC11]. Since these softwares were not in Android Play store, it is difficult for users to know their specification [CJLF13]. This phenomenon and issue of Android system make us believe it is curial to help users make proper decisions about their Android applications based on their own preferences.

2.2 Existing Works about Mitigating Privacy Risks of Mobile Application

There has been a great deal of works on providing privacy to users for smartphone applications. We classify related work into three categories: (1) Privacy detection and analysis (2) Privacy protection and risk mitigation (3) Understanding users' privacy for different applications. After reviewing existing works, we compare some representative research works about smartphone applications privacy in table 2.1.

Privacy detection and analysis

With the proliferation of applications in mobile operating systems, their drawbacks have aroused much public concern. The research community has put much effort on detecting and analyzing the potential privacy risks of smartphone applications. Two mainstream methods are static analysis and dynamic analysis, while some methods are also proposed from different perspectives.

The static methods analyze the source code of smartphone applications to generate a control flow graph (CFG) rather than actually executing the applications. After covering all the paths of CFGs, the methods would detect and analyze the potential privacy risks of each smartphone application. In the most cases, the applications are often detected and analyzed by an automated tool or system. Some such works have been designed. LeakMiner

is tool to detect disclosure of sensitive information on Android based on static analysis [YY12b]. It can identify 145 real leakages in a set consisting of 1750 applications, even though with 160 false positives. Mann and Starostin [MS12] design a framework to detect privacy leakage for Android applications through static information flow analysis. It tried to identify whether the Dalvik bytecode implementation of an Android app conforms to a given privacy policy. AndroidLeaks is a static analysis framework for finding potential leaks of sensitive information in Android applications on a large scale [GCEC12]. It found that there are 57,299 potential privacy leaks in 7,414 applications among 24,350 tested applications. ComDroid has been proposed to help developers to analyze their own applications before release, the custom code has potential privacy risk since the code is usually unjustified [CFGW11]. Some static analysis tools also have been developed to automatically detect attempts to load external code [PFB+14]. Applying the static analysis to Android permission mechanism is also a telling method. Android permission mechanism allows each application has permissions to perform any operations that would adversely impact other applications, the system and users. The permission abuse however also can lead data leakage. Woodpecker is a tool which aims to identify the permissions or capabilities abuse of applications using static analysis [GZWJ12]. It found that 11 permissions were leaked among 13 privileged permissions examined. All aforementioned tools and systems are for Android platform, PiOS is a tool in iOS platform, which allows people to statically analyze applications for potential information disclosure [EKKV11]. According to their findings, it claimed that most applications respect personal identifiable information stored on user's devices in the light of testing of over 1,400 iOS applications.

The static detection and analysis methods have to spend more time on scrutinizing the source code of mobile applications and generating CFG for further analysis, yet they actually have no time performance overhead since processing is done before the applications are launched. Moreover, there is no prerequisite for static analysis, like mobile devices or simulation environment, which are necessary to dynamic detection and analysis.

Unlike the static methods, dynamic detection and analysis methods monitor the applications when they are running. In other words, the actual behaviors of mobile applications would be analyzed. Data flow analysis (DFA) is a technique to achieve the goal by tracking the flow of sensitive data of users [LD14]. However, as we discussed in the static analysis, the dynamic detection and analysis are based on either real mobile devices or simulation environment. The time performance overhead is an obvious drawback.

TaintDroid [EGC+14] is a dynamic taint tacking and analysis system, which involved some aforementioned methods to simultaneously tracking multiple sources of sensitive data. It can provide realtime analysis by leveraging Android's virtualized execution environment. In the architecture of TaintDroid, it predefined nine situations of the information as taint. After monitoring and analyzing the behaviors of 30 third-party Android applications based on the situations, TaintDroid found that 68 instances of potential misuse of users' private information across 20 applications. DroidScope is also proposed as a dynamic analysis platform using virtualization-based malware analysis [YY12a]. It focused on reconstructing both the OS-level and Java-level semantics, by mirroring three aspects of an Android device: hardware, OS and Dalvik Virtual Machine. Dynamic detection and analysis via graphical user interface is another telling method. AndroidRipper can test Android application based on a user-interface driven ripper that explored the application's GUI with the aim of exercising the application in a structured manner [AFT+12].

Currently, more and more detection and analysis methods have emerged besides static and dynamic analysis. It is an intuitive idea to combine these two methods to improve performance of analysis. For instance, Smartdroid is a hybrid static and dynamic analysis method to reveal UI-based trigger conditions in Android applications [ZZD+12]. More specifically, it firstly used the static analysis to extract various expected activity switch paths and then took advantage of the dynamic analysis to scan each UI elements and explored the UI interaction paths towards the sensitive APIs. However, It is resource-consuming when we do such analysis in the smartphone. Thus, Paranoid Android is proposed to address the issue by leveraging cloud-based analysis [PHAB10]. Most analysis works will be finished by

servers rather than smartphones due to the constraints.

Crowdsourcing also has been used for determining vulnerabilities in the smartphone due to its unprecedented ability of data collection. Crowdroid is a framework for collection of traces from an large number of real users based on crowdsourcing mechanism [BZNT11].

Privacy protection and risk mitigation

Privacy protection for smartphone applications is another challenging issue. A lot of research have focused on how to preserve users' privacy and applications' functionality at the same time. We classify the existing works into three groups, (1) permission removal (2) access control (3) data mock.

Android system provides a permission mechanism to protect users' data. The permission list of an application will be shown to users before they install applications from the app store. Only when the applications get approbation can they be installed. After installation, the applications can access the resources and information according to their permission lists. In nature, Android permission mechanism intends to improve users' awareness of the privacy about the applications and preserve the privacy.

However, most Android users have defective understanding about the permissions. Making things worse, they paid limited attention to the permission list, which is shown on the screen just before the installation [KCC+12]. A laboratory study showed that Android users have little attention and comprehension to the applications and corresponding permissions of data usage [FHE+12]. Thus, a feasible way to mitigate data abuse is to establish a system which can prevent applications from accessing resources without the stated permissions [OTK+12]. In this case, users will know what kind of information will be accessed by the app. However, some developers always ask for the unnecessary permissions due to ambiguous API documentations and bad development habits [FCH+11]. This abuse of permissions also lead unexpected information disclosure. An immediate idea is to remove or constrain the suspected permissions. Permissions removal has been proposed to mitigate the privacy leakage in Android smartphones [DMC14]. It is a kind of reverse-engineering process which aims to remove an app's permission to a resource when the permission is unrelated to the

application. The repackaged application can run in the smartphone again. A key challenge is how well-integrated the various permissions are within an application.

Access control provides a different perspective of detecting and protecting privacy in smartphone. FlaskDroid privodes mandatory access control on Android's middleware and kernel layers to prevent information disclosure [BHS13]. All these works focused on analyzing and protecting privacy for Android applications. However, as we discussed before, protecting users' information unilaterally cannot meet their requirements since users have different concerns towards Android applications.

AppIntent provides a framework which attempts to control data transmission to prevent Android applications from stealing sensitive data, meanwhile identified if transmission is from users' intention [YYZ+13]. For each data transmission, it can generate a sequence of GUI manipulations corresponding to the sequence of events. Thus, it can help analysts to determine if the data transmission is user's intention or not. TrustDroid is designed to isolate data and communicated at different layers of Android software stack, including the middleware layer, kernel layer and network layer [BDD+11]. AppFence is a method which aims to empower users to protect their data from exfiltration by permission-hungry applications [HHJ+11]. It can covertly substitute shadow data in place of data that the user wants to keep private, and block network transmissions that contain data the user made available to the application for on-device use only.

However, as suggested by Ghosh et al. [GJFJ12], current privacy control mechanisms are static and cannot preserve privacy in a dynamic context-aware environment. According to recent systematical research, several vulnerabilities have existed in Android applications. Their presence appears even in some extremely popular applications [ZJ13]. Thus, plenty of works focus on privacy of Android platform and its applications. Techniques and tools that can detect and prevent information from being leaked in Android applications have been widely studied [PFB+14, BHS13]. Permission analysis is a telling method to detect sensitive information potential leakage. It can scrutinize Android app to know whether the developers follow least privilege with their permission requests. In the light of the results

of an elaborate analysis, limitations of Android's UID sharing method coerce developers to write custom code rather than rely on OS-level mechanisms [BCMVO12].

Some applications cannot run without accessing specific information so data mock mechanism can address the issue. TISSA [ZZJF11] and MockDroid [BRSS11] can provide artificial data instead of real information to the applications so that they can still function. In this case, there is no risk for users elaborate because the data is fake. However, due to the same reason, applications cannot provide competent services to the users.

All aforementioned works provide us some techniques and tools to detect and mitigate privacy risk on Android applications. However, these works did not tell users about the information to be offered and data to be preserved. Understanding users' privacy concerns therefore is becoming a significant issue.

Understanding users' privacy

Understanding user's privacy is based on human-centric privacy. The privacy is regarded as the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others [Wes68]. Therefore, it should emphasize that users have adequate awareness and understanding to their personal and sensitive information. According to a recent survey [BGR13], Android users hold quite different viewpoints due to their demographic characteristics, privacy awareness, and reported behaviors when installing applications. Users may be surprised when they realize data collection and distribution of smartphone applications [SMSB14].

It is also challenging to recognize users' perceptions of whether a given action is legitimate, or how the action makes them feel with respect to privacy. A model, privacy as expectations, is proposed to capture people's expectations of privacy [LAH+12]. Appprofiler [RQM13] is an approach to provide users with knowledge for decision-making about Android applications through analyzing privacy-related behaviors of the applications and the users' opinions. After understanding users' perceptions, it is important to assess privacy risks and predict users' privacy preferences. An approach is proposed for assessing the privacy risk of Android users based on the impact valuation from users and their profiles

[MTG13]. Super-Ego [Toc14] is a crowdsourcing framework which can predict the users' privacy preferences for different locations on the basis of the general user population. Altman's theory of boundary regulation [Alt77] and Nissenbaum's theory of contextual integrity [Nis04] are also adopted to explore the privacy gap between users' privacy expectations and smart phone usage.

Table 2.1: Comparison of some representative research work about smartphone applications privacy

Methods/Systems	Privacy Characteristics	Objectives	Platform	Summary
Woodpecker [GZWJ12]	Technology-centric	Privacy detection and analysis	Android	13 privileged permissions were examined and 11 were leaked, with individual phones leaking up to eight permissions.
TaintDroid [EGC+14]	Technology-centric	Privacy detection and analysis	Android	30 popular Android applications were examined, 68 instances of potential misuse of users' privacy were found across 20 applications.
AndroidLeaks [GCEC12]	Technology-centric	Privacy detection and analysis	Android	24,350 Android apps were examined, 57,299 potential privacy leaks in 7,414 Android applications were found
PiOS [EKKV11]	Technology-centric	Privacy detection and analysis	iOS	More than 1,400 iOS apps were analyzed, 656 Apps use ad library code which may disclose devices ID, 36 Apps leak GPS location and 5 Apps leak contacts.
LeakMiner [YY12b]	Technology-centric	Privacy detection and analysis	Android	It is an automatic and static taint analysis method. After analyzing 1750 apps, it can identify 145 real leakages in this app set.
ComDroid [CFGW11]	Technology-centric	Privacy detection and analysis	Android	It can be used by developers to analyze their applications before release. It analyzed 20 applications and found 34 exploitable vulnerabilities; 12 of the 20 applications have at least one vulnerability.

FlaskDroid [BHS13]	Technology-centric	Privacy protection	Android	It provides mandatory access control simultaneously on both Android's middleware and kernel layers. The evaluation is based on the empirical testing using the security models, testbed of known malware and synthetic attacks.
TrustDroid [BDD+11]	Technology-centric	Privacy protection	Android	It is a framework, which can isolate data and applications at different layers (middleware layer, kernel layer, network layer) with a negligible overhead, small cost on battery's life-time
MockDroid [BRSS11]	Technology-centric Human-centric	Privacy protection	Android	It is a modified version of the Android which allows a user to provide artificial data to the apps such that they can still function (possibly with reduced functionality). A random sample of twenty-three popular applications can successfully run in the MockDroid.
AppIntent [YYZ+13]	Human-centric	Privacy protection, Privacy detection and analysis	Android	It is an analysis framework, which can provide a sequence of GUI manipulations corresponding to the sequence of events to determine if the data transmission is based on users' intention or not. The evaluation is based on a set of 750 malicious apps and 1,000 top free apps from Google Play.
Privacy as expectations [LAH+12]	Human-centric	Understanding users' privacy	Android	It is a system aims to capture users' expectations of what sensitive resources mobile apps use through crowdsourcing. It found that both users' expectation and the purpose of sensitive resources can affect users' feelings and their decisions. It also claimed that proper notification about the purpose of resource access can ease users' privacy concerns to some extent.

ProtectMyPrivacy [AH13]	Human-centric	Understanding users' privacy	iOS	More then nine months, ProtectMyPrivacy were used by more than 90000 real users. It also reviewed more than 200000 applications and recommend users about the decision for their permission settings in iOS.
Appprofiler [RQM13]	Human-centric	Understanding users' privacy	Android	It provides users the knowledge needed to make informed decisions about the applications they install. There are three findings: 1) the permission mechanism is not very fine-grained; 2) it is important to differentiate between actions performed by users and by the system; 3) There is a huge gap between third-part library code and the code in a specific application.

2.3 Existing Works about Preserving Privacy in Mobile Participatory Sensing

In this section, we review the related works about smartphone privacy in mobile participatory sensing. In the mobile participatory sensing, smartphone users need to provide sensory data. By comparison with other data communication in smartphone, mobile sensing has more possibilities to access users' information legally. We therefore study some works which devoted to preserve privacy in mobile sensing.

For better understanding of the existing works, we firstly provide an overview of system model of mobile sensing. The typical architecture of mobile sensing system is illustrated in Fig. 2.1. Normally, there are two significant roles in mobile sensing applications, participants and stakeholders [LML+10]. The participants refer to the people who accept the sensing tasks, and collect the data from physical world. The participants capture the data using different mobile devices, while our study focuses on smartphones. Stakeholders refer to the people who benefit from the data. They usually initiate a mobile sensing application

and then access the sensory data collected by participants for further analysis or presentation. A number of participatory sensing applications have emerged recently [LML+10]. However, users' privacy concern is an obstacle for long-term deployment [ABK12]. Based on the characteristics of smartphone privacy, we classify related work into two categories: (1) Privacy preservation and awareness (2) Personalized privacy risk mitigation.

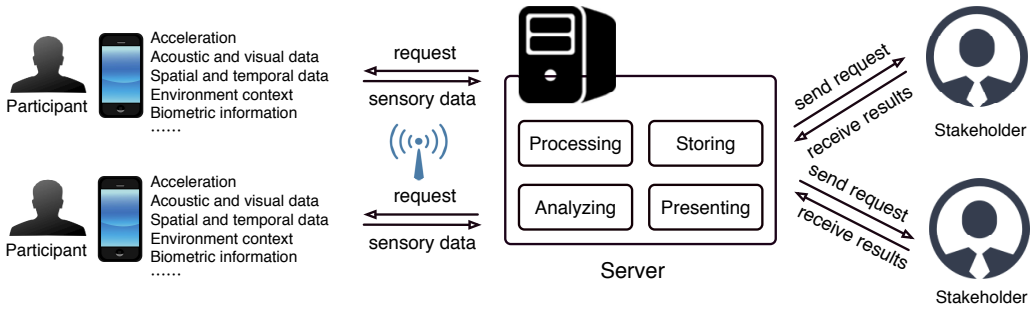


Fig. 2.1: The typical architecture of mobile sensing system

Privacy preservation and awareness

Privacy preserving is a long-standing issue and prompts a wide discussion in mobile sensing systems, especially for the systems which aims to the pervasive information collection. As we mentioned before, privacy issue may be a cardinal obstacle [GCJW10]. Many works on privacy preserving has been proposed.

PiRi, a privacy-aware framework for participatory sensing systems, which attempted to address the privacy issues based on an untrusted central data server model and enabled participation of the users without compromising their privacy [KS11b]. PEPSI, a privacy-enhanced participatory sensing infrastructure, explored realistic architectural assumptions and a minimal set of formal requirements aiming at protecting the privacy of both data producers and consumers with low additional cost and overhead [DCS11]. AnonySense is a privacy-aware architecture for realizing pervasive applications based on the participatory sensing by mobile devices [SCP+11]. It allowed the applications to submit various sensing

tasks. The tasks are distributed across anonymous participating mobile devices, later receiving verified, yet anonymized, then sensory data are collected from the environment. SensorSafe is an architecture for managing personal sensory information in a privacy-preserving way, which consists of multiple remote data stores and a broker [CCCS11]. Thus, the users can retain the ownership of their data, the management of multiple users can be supported as well. It provided a context-aware fine-grained access control mechanism by which users can determine their own sharing rules based on various conditions including context and behavioral status. Jedrzejczyk et al. observed anonymous users of a location-based social networking application in their natural environment and demonstrated how to re-identify them based on the data [JPBN09].

Furthermore, a privacy-preserving mobile sensing scheme for multidimensional data which used negative surveys has been presented. In this scheme, the server can reconstruct the probability density functions of the original distributions of sensory values, without knowing the participants' actual data [GEH+12]. Location privacy is another important concern in mobile sensing systems [Min04]. A decentralized mechanism to preserve location privacy during the collection of sensor readings has been proposed, which can exchange the sensor readings of users in physical proximity for jumbling the location information [CGR+11]. Another algorithm is designed to address kNN queries for datasets which grouped by users based on locality-sensitive hashing in mobile sensing systems [VZG12]. User-side privacy-protection scheme can adaptively adjust the parameters of participatory sensing for satisfying individual location privacy protection requirements against adversaries in a measurable manner [APN+14].

Personalized privacy risk mitigation

Unlike privacy preserving mechanism, personalized privacy methods consider users' requirements in mobile sensing, which served as the underpinning of further protection. Muslukhov et al. held the viewpoint that users' privacy requirements in mobile sensing systems are heterogeneous, and different targets and form of utilizing personal data can affect users' concerns [MBK+12]. For example, Barkhuus and Dey presented an experimental case study

to examine people's concerns for location privacy [BD03]. They found that location tracking services can cause more concerns for privacy than position-aware services. Gedik and Liu provided a privacy personalization framework to support location k-anonymity for context-sensitive personalized privacy requirements [GL08]. Each mobile node is specified with the desired minimum level of anonymity and maximum temporal and spatial resolutions.

A suite of scalable spatio-temporal cloaking algorithms, CliqueCloak, is designed to avoid or reduce known location privacy threats before forwarding requests. Gong et al. proposed a dynamic privacy management system, which is capable of enabling tangible privacy control and feedback in a pervasive sensor network [GLP10]. A key contribution in this work is to conduct a user study for showing some insights of privacy/benefit tradeoff from various sensing capabilities and reflect how privacy settings and user behavior relate. For instance, Freudiger et al. extended the understanding of the privacy risk in the location-based services [FSH12].

Table 2.2: Comparison of some representative research work about privacy in mobile sensing

Methods/Systems	Privacy Characteristics	Objectives	Summary
AnonySense [CKK+08]	Technology-centric	Privacy preserving	It allows the sensing tasks distribute anonymously to the participating mobile devices. The sensory data will be reported back in a verified and anonymized way as well.
Sensorsafe [CCCS11]	Technology-centric	Privacy preserving	It is a system for managing personal sensory information in a privacy-preserving way with supporting multiple users. Users can define their sharing rules based on different context through access control mechanism.
PEPSI [DCS11]	Technology-centric	Privacy preserving	The main contribution is the work is based on some realistic assumptions and a minimal set of formal requirements aiming at protecting privacy of both data producers and consumers. Meanwhile, adding low computational cost and communication overhead is another contribution.

PiRi [KS11a]	Technology-centric	Privacy preserving	PiRi is a privacy-aware framework, which aims to guarantee predefined users' privacy when they participate in mobile sensing system. Thus, defining privacy is one of the main contribution.
Prisense [SZL10]	Technology-centric	Privacy preserving	It is a privacy-preserving data aggregation method based on data slicing, data mixing and non-additive aggregation functions to against a tuneable threshold number of colluding users and aggregation servers.
PoolView [GPTA08]	Technology-centric	Privacy preserving	It provides privacy guarantees on stream data for participatory sensing application, which is in the light of data perturbation and reconstruction techniques. The actual data is applied in the evaluation for demonstrating the privacy-preserving aggregation functionality.
Adaptive Personalized Privacy [APN+14]	Human-centric	Personalized privacy	It considers heterogeneous user privacy requirements in mobile sensing system. A user-side privacy protection adaptively adjusts parameters to meet personalized privacy is proposed, which attempts to balance the privacy and utility. The evaluation is based on synthetic and real data.
CliqueCloak [GL05]	Human-centric	Personalized privacy	It provides a privacy personalization framework to support location k-anonymity for context-sensitive personalized privacy requirements. Each mobile node is specified the desired minimun level of anonymity and maximum temporal and spatial resolutions. Its main contribution is to avoid or reduce known location privacy threats before forwarding requests.
Dynamic Privacy Management [GLP10]	Human-centric	Personalized privacy	It is a dynamic privacy management system aimed at enabling tangible privacy control and feedback in a pervasive sensor network. A key contribution is to conduct a user study to show some insight of privacy/benefit tradeoff from various sensing capabilities and how privacy settings and user behavior related.

Chapter 3

Privacy Measurement based on Users' Preferences towards Data Sharing in Mobile Participatory Sensing Systems

In this chapter, we investigate the privacy measurement in mobile participatory sensing. We propose a human-centric privacy measurement method, PriMe, which can quantify the privacy risk based on users' preferences towards data sharing in mobile participatory sensing systems.

This chapter is organized as follows: Section 3.1 is the overview of this work. We illuminate system model of participatory sensing in Section 3.2 and show the privacy in participatory sensing in Section 3.3. Subsequently, we elaborate on PriMe in Section 3.4 and demonstrate experiments and results in Section 3.6. Finally, Section 3.7 concludes this chapter.

3.1 Overview

Ubiquitous and increasingly capable mobile devices bring forth so-called mobile participatory sensing systems. The idea behind these systems is that individuals and communities use mobile deceives to collect, analyze, and share data regarding their environments for use

in discovery. Many such mobile participatory sensing systems have been developed over the years, and some also deployed in the real world [KXAA13]. One of the main obstacles for a long-term real-world deployment of such systems is privacy issues. Privacy in a participatory sensing system has particular characteristics. On the one hand, users have to provide their data in order to participate and keep the system running. On the other hand, users are generally ambivalent when it comes to sharing any information due to privacy concerns [LCYZ15]. Some works on preserving privacy in participatory sensing systems have been published in recent years, e.g., [CRKH11]. However, we argue that privacy is not a static concept, but rather fluid and malleable as the perception of privacy differs from person to person. Users need to understand the implications of the data they are supposed to share regarding their personal privacy in order to make an informed decision about participating in sensing tasks or not. However, assessing the risk to one's personal privacy for every sensing task is very arduous. In order to automate this process, it is necessary to understand and model a user's privacy risk with regard to their personal perception.

In this work, we propose the human-centric privacy measurement method *PriMe*. To the best of our knowledge, it is the first privacy measurement method for mobile participatory sensing systems that is based on the user's perception. For each sensing task, PriMe quantifies the privacy risks for each user individually based on his/her preferences towards sharing certain types of data. For this, we propose two intuitive properties of user preferences and regard them as metrics: 1) intrinsic sensitivity, i.e., the individual inherent sensitivity, and 2) extrinsic sensitivity, i.e., a person's sensitivity towards different data items due to data features. Then, we determine each users privacy risk by by quantifying and aggregating these two properties. To prove our proposed method, we implemented, deployed, and evaluated PriMe with real world users (65 recruited volunteers from different backgrounds). The results show that PriMe is able to make accurate measurements that satisfy users, and thus

are widely accepted as a trustworthy tool.

3.2 System Model of Participatory Sensing

We first provide an overview of the system model of typical participatory sensing systems. Subsequently, we elaborate on the procedures, tasking and sensing.

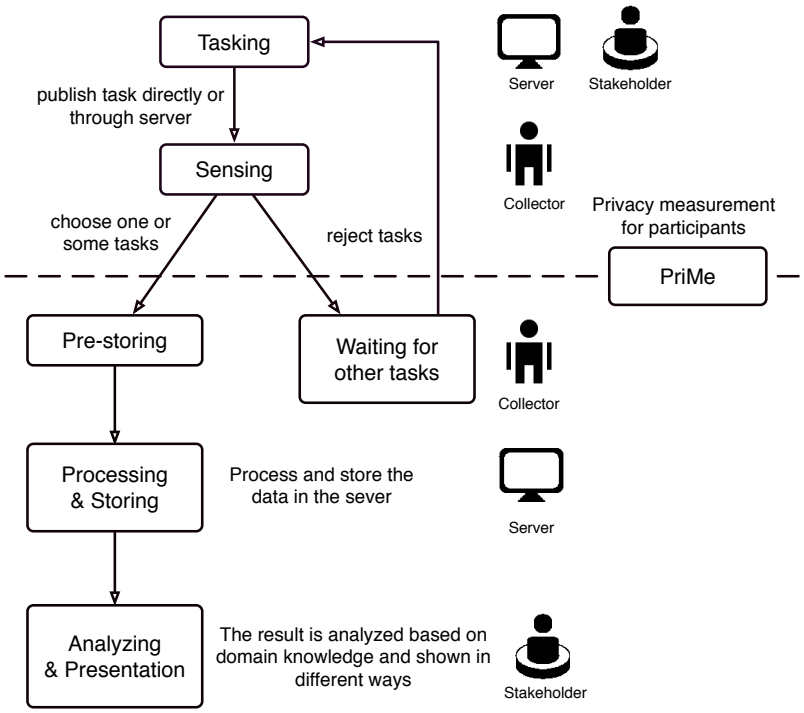


Fig. 3.1: In typical architecture of participatory sensing, PriMe quantifies the privacy for each user in participatory sensing

Numerous participatory sensing applications have been emerged recently [CLCC15]. The typical architecture of participatory sensing is shown in Fig. 3.1. Normally, there are two significant roles in participatory sensing systems, participants and stakeholders. Participants refer to the people who participate in sensing task and capture the multifarious data from physical world. Stakeholders refer to the people who benefit from the data. They usually initiate a participatory sensing application, and then integrate or analyze the sensory data. More specifically, the participants capture the data from the environment using

different mobile devices. Afterwards, the data is pre-processed and transferred to the server. The server will process the data and transfer the result to different stakeholders for further analysis.

The sensing tasks can be one-off (e.g., monitoring noise around your home this morning) or repeating (e.g., reporting health condition daily). The stakeholders published sensing tasks to the participants through the server or pushes the tasks to the participants directly [Kan11]. Participants may receive many tasks from different participatory sensing applications. They can choose some of tasks to participate according to their interests. In practise, sensing tasks can be assigned to the participants, such as monitoring traffic information, capturing PM2.5 of a specific area, collecting their biometric record and so forth.

The high quality and accuracy of the data can only be achieved when sufficient number of participants is involved in the participatory sensing systems. Hence, a sensing task should be assigned to a group of participants called sensing group. A sensing group is a number of participants who join in and complete a task in participatory sensing task. A participant can be a member in several sensing groups, when he/she participates in different tasks at the same time. For a participatory sensing task, a participant need to provide a variety of data, including personal information and sensory data. In the light of ever-more-capable mobile devices, participants can capture multitudinal data from the physical world easily. The time [CRH+13], location [JVLL12], acoustic and visual data [LPA+11], acceleration [VKSW11], environmental context [EC12] and biometric information [CS13] are collected for different tasks.

3.3 Privacy in Participatory Sensing

Our work presented in this work provides a privacy measurement method based on users' preferences that enables users to better understand their privacy risks in participatory

sensing systems. For this, we first clarify the semantic of privacy in participatory sensing systems.

We refer to two acknowledged privacy definitions as proposed in the literature: *"the right to be alone due to private life, habits, acts, and relations"* [WB90] and *"the claim of individuals, groups or institutions to determine for themselves when, how, and to what extent information about them is communicated to others"* [Wes68]. The definitions emphasize that privacy is the ability of an individual or a community to seclude themselves or conceal information about themselves from others. In effect, privacy should be driven by individual preference, instead of being one absolute for all. The same holds true in participatory sensing systems. Therefore, we believe that *privacy in participatory sensing depends on the participants' sensitivities towards the data in question.* We consider two types of sensitivity metric:

1. *intrinsic sensitivity*; This metric indicates the human nature about privacy. However, people's privacy concerns differ from person to person. Some people inherently have a higher/lower sensitivity than others. For example, one person may really care about his/her contacts, whereas others may not care about this information at all.

2. *extrinsic sensitivity*; This metric illustrates the data feature about privacy. Some data are more sensitive than others. For instance, location naturally can disclose more information than the data about your favourite food, no matter what kind of people you are. Extrinsic sensitivity also shows a person's sensitivity may vary with different data and scenarios. For example, a user may care less about his/her weight information in discussion with a health mentor, even though the weight information is very sensitive based on his/her intrinsic sensitivity.

Thus, the two types of sensitivity metric describe subjective and objective factors in a sense. That is the reason why we need to consider both of them. Based on this, PriMe

measures each users' privacy risk by quantifying and combining the intrinsic and extrinsic sensitivities of the information in question. Next, we present our approach.

3.4 Privacy Measurement

We propose PriMe, which can quantify a participant's privacy. The result of PriMe's measurement is a floating point number from zero to one, with a higher number indicating a higher privacy risk. In this section, we elaborate the quantification of privacy by formalizing it in the context of a participatory sensing system.

We assume that there are N participants in a participatory sensing application. Each participant has L data items. Participants can set a privacy tag to each data item to present their willingness to share information associated with the data item. All privacy tags for participants form the $N \times L$ matrix M. At this moment, participants set their privacy tags by use of the dichotomous variable $\{0, 1\}$. More specifically, 0 denotes that the participant does not want to share that datum with anyone, whereas 1 means the participant allows the disclosure of that information. The rows of M correspond to participants, and the columns of M correspond to profile items, respectively. $M_{i,j} = 0$ represents that the profile item j of participant i is private, whereas $M_{i,j} = 1$ denotes the data item is public and can be shared with others. The following two examples further illustrate the situation.

Example 1: In a participatory sensing application, a participant i has three data items, $j_1 = \{email\,address\}$, $j_2 = \{current\,location\}$ and $j_3 = \{age\}$. Thus, $M_{i,j_3} = 1$ means that participant i is willing to share his/her age, and $M_{i,j_1} = 0$ indicates that the participant is reluctant to provide his/her email address. Disclosing the current location or the email address of a user is usually more threatening than revealing a user's age, because the location or the email address – which we then call *sensitive information* – can identify the exact person, whereas age cannot.

Example 2: Another participant i' has the same data items, $j_1 = \{email\,address\}$, $j_2 = \{current\,location\}$ and $j_3 = \{age\}$. However, the participant sets $M_{i,j_1} = 0$ and $M_{i,j_2} = 1$, which means that he/she cares greatly about the email address rather than current location. In this case, location is not a sensitive information any more, even though it is very sensitive to many other people. That is, this particular participant does not feel his/her privacy threatened, if the location information is leaked.

In light of these examples, it is easy to see that a person's privacy preference is a crucial property to determine the privacy threat. Thus, we consider the property of a participant's individual sensitivity (i.e., a participant's privacy preference) rather than the sensitivity of each data because the former is a more inherent property than the latter. Individual sensitivity is an inborn property and can be shaped by a long-term implication of the environment. However, the sensitivity of each data item differs from person to person, scenario to scenario. The general way to identify such sensitivity is to detect from a large sample of people but it still rely on a person's individual attitude towards certain data. According to the two properties, we define $\delta_{i,j}$ as participant i's extrinsic privacy preference of item j and β_i as the intrinsic sensitivity of participant i. $Pr_{i,j}$ denotes the privacy risk of information item j of participant i when item j is provided. Based on these two parameters, we quantify users' privacy by drawing inspiration from the Rasch Model [BF13], as shown in Equation 3.1.

Before we continue presenting our approach, we first discuss why we chose the Rasch Model to quantify the privacy risk in participatory sensing.

$$Pr_{i,j} = \frac{e^{\beta_i - \delta_{i,j}}}{1 + e^{\beta_i - \delta_{i,j}}} \tag{3.1}$$

The Rasch Model is a psychometric model for measuring/analyzing categorical data as a function of the trade-off between (a) the respondent's abilities and attitudes, and (b) the item difficulty to a particular respondent. A typical application of the Rasch Model is, for

example, to estimate the probability of people answering questions correctly based on the ability of a person and the hardness of the question as perceived by the person.

We think the relationship between a users' privacy and their respective attitude towards each item fits this model. More specifically, based on willingness a participant is to reveal his/her information, we also can estimate the probability of a user perceiving a data as sensitive, which can be regarded as a privacy measurement. In the Rasch Model, β_i represents the ability of person i and $\delta_{i,j}$ denotes the difficulty of each question to a specific person, and the result is the probability of a correct response to a given assignment. In our scenario, we can map the two parameters exactly to the sensitivity metrics we describes previously. Thus, there are two parameters, β_i and $\delta_{i,j}$, that need to be computed.

Next, we show how to estimate β and δ based on the matrix M. For this, the maximum-likelihood estimation (MLE) method can be used because this method maximizes the likelihood, or the probability, of our observation and thus is naive and should be the first to think of.

Before we step into the likelihood function, we look deep into the data first. Any participant's decision will not affect the others. Thus the tags are independent across participants. For each participant, the j items can be grouped based on their relative sensitivity to the participant. Some items are similar to a participant so he/she would have the same probability to reveal information on these items. In our setting, $\delta_{i,j}$ would be the same. Thus we can believe that $Pr_{i,j}$ would be the same in one group. The classification of groups need not be the same for different participants. Suppose the j items are classified into G_i groups for participant i and any item falls into a group $g_{ik}, k = 1, 2, \cdots, G_i$. Choices among different groups should be independent for the same participant. Choices within groups should also be independent but identically follow a Bernoulli distribution with parameter

$Pr_{i,g_{ik}}$. Therefore, the likelihood function for M is

$$\mathcal{L}(\beta, \delta|M) = \prod_{i=1}^{N} \prod_{k=1}^{G_i} \prod_{j \in g_{ik}} Pr_{i,g_{ik}}^{M_{i,j}} (1 - Pr_{i,g_{ik}})^{(1-M_{i,j})} \tag{3.2}$$

where $Pr_{i,g_{ik}} = Pr(M_{i,j}|\beta_i, \delta_{i,g_{ik}})$. The estimators are the ones that maximize the above likelihood function, i.e.,

$$(\hat{\beta}_i, \hat{\delta}_{i,g_{ik}}) = \arg \max_{\beta_i, \delta_{i,g_{ik}}} \mathcal{L}(M_{i,j}|\beta_i, \delta_{i,g_{ik}})$$

The values can be achieved simply by taking derivatives of the above likelihood function with respect to the two parameters β and δ, and then setting them to zero. Since the logarithm function is monotonically increasing, we can take the derivatives of the logarithm of the likelihood function and then set them to zero. We follow the steps for each $i = 1, 2, \cdots, N$ and $j = 1, 2, \cdots, L$.

By use of Equation 1, we finally get

$$(1 - Pr_{i,g_{ik}}) \sum_{j \in g_{ik}} M_{i,j} - Pr_{i,g_{ik}}(|g_{ik}| - \sum_{j \in g_{ik}} M_{i,j}) = 0 \quad \text{for } \forall k = 1, \cdots, G_1 \tag{3.3}$$

where $|g_{ik}|$ is the number of elements in group g_{ik}. Therefore, we achieve

$$Pr_{i,g_{ik}} = \frac{e^{\beta_i - \delta_{i,g_{ik}}}}{1 + e^{\beta_i - \delta_{i,g_{ik}}}} = \frac{1}{|g_{ik}|} \sum_{j \in g_{ik}} M_{i,j} = \bar{M}_{i,g_{ik}} \tag{3.4}$$

That is,

$$\beta_i - \delta_{i,g_{ik}} = \log \frac{\bar{M}_{i,g_{ik}}}{1 - \bar{M}_{i,g_{ik}}}. \tag{3.5}$$

Equation 3.4 makes sense based on our premises. From our settings, participants inherently make the same decisions on items in the same group. The inherent probability that participant i will reveal information on items in this group g_{ik} is $Pr_{i,g_{ik}}$, which can therefore be estimated by the average values of tags in this group.

We only get one equation but we have two unknown parameters. Therefore, we need to seek other methods to estimate the parameters. From Equation 3.5 we can see that if we know either $\vec{\beta}$ or $\vec{\delta}$ then we will know the other one. Yet it is easy to see that if $\vec{\delta}$ is given then $\vec{\beta}$ will be simpler to solve. Thus we do the estimation by iteration starting from estimating $\vec{\beta}$ given $\vec{\delta}$.

First, we give some initial values to $\vec{\delta_i}$, and then estimate β_i using the Bayesian method. When we get the posterior probability of β_i, it is trivial to estimate β_i by its mode. By standard convention [BF13], β_i would have a Gaussian prior distribution, with some mean μ and variance σ^2. The posterior probability for β_i is

$$\begin{aligned}
&\mathbb{P}(\beta_i | M_{i,j}, j = 1, \cdots, L, \vec{\delta_i}) \\
&= \frac{\mathbb{P}(M_{i,j}, j = 1, \cdots, L | \beta_i, \vec{\delta_i}) f(\beta_i)}{\int \mathbb{P}(M_{i,j}, j = 1, \cdots, L | \beta_i, \vec{\delta_i}) f(\beta_i) d\beta_i} \\
&\propto \mathbb{P}(M_{i,j}, j = 1, \cdots, L | \beta_i, \vec{\delta_i}) f(\beta_i) \\
&\propto \prod_{k=1}^{G_i} \prod_{j \in g_{ik}} \frac{e^{\beta_i - \delta_{i,j}}}{1 + e^{\beta_i - \delta_{i,j}}} e^{-\frac{(\beta_i - \mu)^2}{2\sigma^2}}
\end{aligned} \tag{3.6}$$

Thus, the estimated β_i is

$$\hat{\beta}_i = \arg\max_{\beta_i} \prod_{k=1}^{G_i} \left(\frac{e^{\beta_i - \delta_{i,g_{ik}}}}{1 + e^{\beta_i - \delta_{i,g_{ik}}}}\right)^{|g_{ik}|} e^{-\frac{(\beta_i - \mu)^2}{2\sigma^2}}. \tag{3.7}$$

Plugging the result into Equation 3.5, we can update the estimated $\hat{\delta}_{i,g_{ik}}$ and then begin our iteration until we converge. Thus, the whole procedure of privacy measurement is shown in Algorithm 1. After successfully estimating the parameters β and δ, it is trivial to quantify the individual privacy risk of a data item.

3.5 Study Methodology

In this section, we describe the study methodology of our work. First, we present the implementation of PriMe based on the system design, from the App and the server side, respectively. Then, we describe the study procedure to demonstrate our experiments.

Algorithm 1: Privacy measurement in participatory sensing systems

Input: Dichotomous matrix M
Output: $\hat{\delta}, \hat{\beta}, Pr_{i,j}$

for $i = 1 \;\; to \;\; N$ **do**
  Classify the L items into G_i groups;
  for $k = 1 \;\; to \;\; G_i$ **do**
    $\delta_{i,g_{ik}}$ = initial_value;
  end
  $\vec{\delta} = \{\delta_{i,1}, \delta_{i,2}, \cdots, \delta_{i,L}\}$;
  while *convergence* **do**
    Calculate $\hat{\beta}_i$ using Equation 3.7;
    for $k = 1 \;\; to \;\; G_i$ **do**
      Calculate $\delta_{i,g_{ik}}$ using Equation 3.5;
    end
  end
  for $j = 1 \;\; to \;\; L$ **do**
    Calculate $Pr_{i,j}$ using Equation 3.1;
  end
end

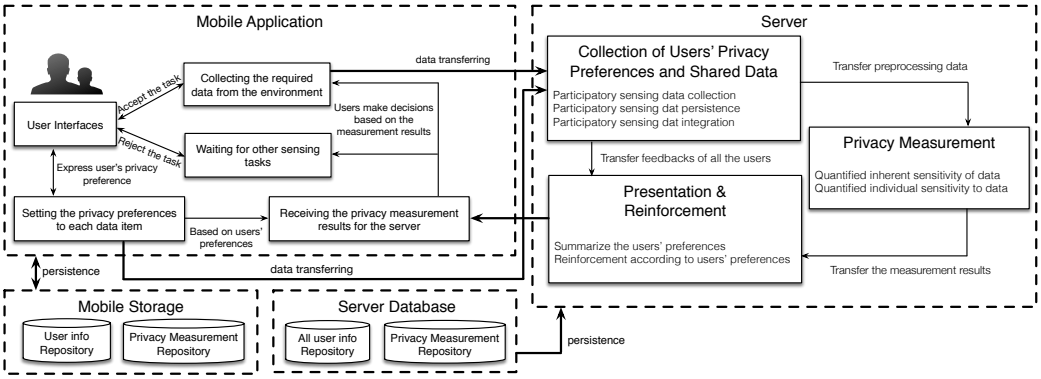


Fig. 3.2: The overview of architecture of PriMe, including the mobile app and the server.

3.5.1 System Implementation

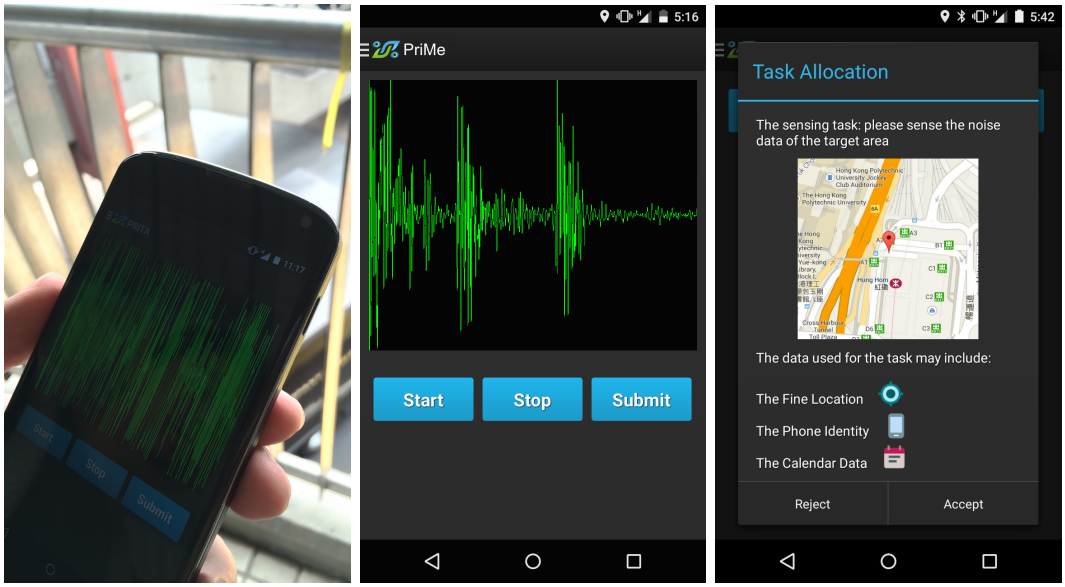
To evaluate our proposed method, we conduct a user study, in which we focus on environmental noise monitoring as the participatory sensing application. The implementation architecture of the system for the user study is depicted in Fig. 3.2. The system consists of an App part and a server part. Throughout this chapter, we refer to this App part as the *PriMe App*. In the study, the participants are required to install the Android application,

to provide them with the participatory sensing function and collect their preferences and feedback towards the privacy assessment of each shared data.

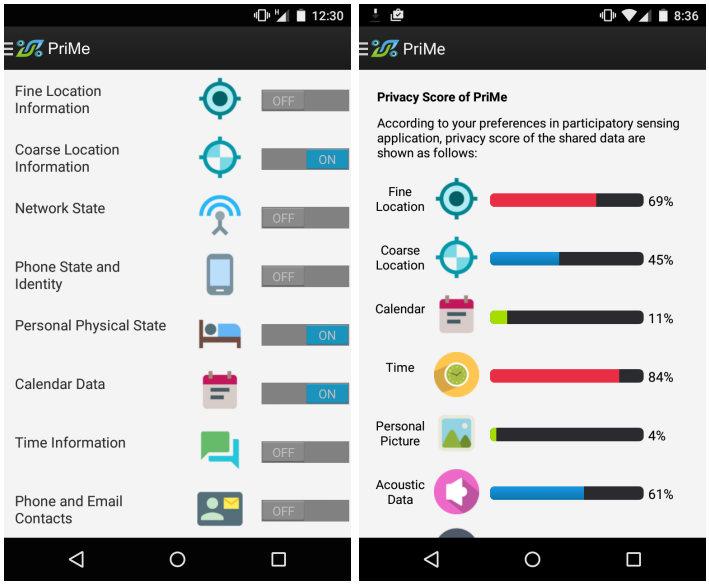
The PriMe App is developed on Android platform. Our prototype of the application is implemented on Android 4.0.3 - 5.1.1 and runs on the Google/LG Nexus 4 handset. There are two design premises of the PriMe App. First, it is a participatory sensing application in nature, so it should receive the sensing tasks and allow the users to provide their collected data. Second, the PriMe App also allows users to express their preferences towards data sharing in the mobile participatory sensing system. According to the premises, we implement and deploy the PriMe App. The key functions are depicted in Fig. 3.3, which is composed by snapshots of the PriMe App on a Nexus 4 phone.

Fig. 3.3(a) and Fig. 3.3(b) show example environment noise sensing acts. In this example, the users make use of PriMe to collect noise data and provide it to the server. They also consider to reject the task due to privacy concerns, as this task requires fine-grained location and calendar information – next to the collected noise data itself. Fig. 3.3(c) shows the notification to a user when he/she is assigned a new task. The description of the task and the required data are listed in these notifications. Fig. 3.3(d) shows that PriMe provides an interface for users to express their preferences towards different data items by switching buttons. In the spirit of participatory sensing, users should not set all data types as sensitive, but only those, which they really care about. According to the users' preferences, PriMe then quantifies the privacy of each users' data, as shown in Fig. 3.3(e).

The server side is designed to receive the collected data from the users and implement our privacy measurement algorithm due to the resource restrictions of mobile devices. As shown in Fig. 3.2, there are three key components in the server. In the collection part, the server mainly focuses on cleaning and structuring the collected data. The privacy measurement is



(a) (b) (c)



(d) (e)

Fig. 3.3: Screenshots of the PriMe App on a Nexus 4. The example participatory sensing application in the study is to monitor the noise in Hong Kong. (a) Participants use the App to monitor noise from the environment. (b) In this case, the sensing data is the environment's noise level for a specific time period. (c) Sensing tasks are assigned to participants through the App, including target area and required data. (d) PriMe provides a user interface for participants to choose their privacy preference for each data item. (e) PriMe details the privacy scores of different data items, with higher scores indicating a higher sensitivity of the user towards the data (the original result is from 0 to 1, we present it in percentage).

based on the collected user preferences towards data sharing. In the presentation and reinforcement part, the server can summarize the collected data and revise the measurements' results based on updated user preferences.

3.5.2 Study Procedure

We deploy PriMe in the real world in order to study its applicability and perception by its users. For the study, we recruited 65 volunteers for the duration of three weeks. As described in the previous sections, the sensing tasks for monitoring the noise levels in specific areas of Hong Kong are assigned to these participants.

The number of participants is determined according to various influential existing works, e.g., [KSK14] and [GPA+10]. Even though our study's size was adequate to evaluate our proposed method with statistical significance, we still plan on conducting a larger study by recruiting participants online in the future. The participants in our user study are from The Hong Kong Polytechnic University, either part time students, full time students, or faculty members. In order to avoid statistical bias and make our results trustworthy, the participants were selected from different backgrounds, genders, and age groups. More specifically, 22 participants are full time students at the university, from various departments. 20 are part-time students, who are also employees in different industries. The remaining 23 participants are faculty members at the university, also from different areas. Details about the participants are shown in Table 3.1.

During the user study, the noise monitoring tasks were randomly assigned to the participants, who were then asked to record the noise signal and upload the data to the server. Further, we asked them to share additional personal information in order to simulate various participatory sensing applications besides noise monitoring. The participant can also decline the task due to their privacy concerns. We did not consider the underlying effect of the decline since we assume the decline is caused by participants' sensitivity. Once PriMe

Table 3.1: Statistics about the Participants in the Study

Category	Participants	Amount	Percentage
Gender	Male	43	66.2%
	Female	22	33.8%
Age	10-19	1	1.5%
	20-24	15	23.1%
	25-29	22	33.8%
	30-40	20	30.8%
	40+	7	10.8%
Background	Energy	3	4.6%
	Materials	1	1.5%
	Industrials	6	9.2%
	Consumer Discretionary	4	6.2%
	Consumer Staples	4	6.2%
	Health Care	8	12.3%
	Finance	9	13.8%
	IT in Security & Privacy	9	13.8%
	IT(except Security & Privacy)	13	20%
	Tele Services	5	7.7%
	Utilities	3	4.6%
Time users spent on smartphones	Rarely (0 1hr)	4	6%
	Sometimes (1 2hr)	14	21.5%
	Frequently (2 4 hr)	27	41.5%
	Very often (4+ hr)	20	31%
Attitudes towards study	Seriously completed	43	66%
	Normally completed	20	31%
	Hastily completed	2	3%

measures a participant's privacy, the server does not send a task which requires highly sensitive data, as perceived by this participant, to this participant, in order to minimize declines. For compiling a list of the most interesting data in terms of frequent use in participatory sensing applications as well as their privacy issues, we studied articles on the Internet, e.g., [mos13], research papers on privacy in participatory sensing, e.g., [FSH12, CRKH11], as well as tips for security and privacy from official guidelines. Table 3.2 lists the resulting set of data, and describes their respective potential privacy risks. In the study, we showed these different data types to the participants in the PriMe App (see Fig. 3.3(d) and 3.3(e)), and asked them to express their respective sensitivities.

At the end of the study, the participants were given a final questionnaire to ask for their feedback. Next, we discuss the results of the study as well as the questionnaire.

3.6 Findings

In this section, we discuss the results of our study, such as the performance of the proposed method, as well as other interesting findings.

3.6.1 Participant Sensitivities

During the study period of three weeks, we not only delivered sensing tasks to the participants, but also asked them to set their preferences towards sharing certain data types. Fig. 3.4 shows the sensitivity measurement results of an example participant. The higher the score, the more sensitive the participant is to the data type in question. For example, we can see that the participant considers fine grained location information as well as calendar information as the most sensitive. Meanwhile, Fig. 3.5 shows a plot of the sensitivity of all participants to a specific data type – coarse location information in this instance. It becomes clear that the individual sensitivities of the participants regarding the same data type may differ greatly. For example, participant no. 32 has the lowest sensitivity towards

Table 3.2: Overview of the most collected data in participatory sensing and their potential privacy risks.

Data Type	Description
Time	Some participatory sensing applications require current time, the format can be shown as 09183302202015. This data will identify temporal information and disclose privacy when other data is combined.
Location	Some participatory sensing applications require current location information, fine-grained location is provided by GPS, coarse-grained location is provided by WiFi or the cellular network. These information will reveal a user's location.
Picture & Video	Pictures and videos are also asked by participatory sensing applications, like taking photos of consumed meals and recording a short video with your family. The content of contributed pictures and collected videos also can reveal personal information about the participants and their environment.
Sound	Sound signals can be captured by smartphones for participatory sensing applications. Given a participant's sound signal, it may allow third parties to determine his/her current context.
Acceleration	Acceleration data is recorded intentionally or automatically during participatory sensing tasks. The data may appear less threatening, but it always can show some clues to leak a participant's privacy.
Environmental Data	Environmental data is often collected since a lot of participatory sensing applications focus on the environment. All the environmental data can indicate spatio-temporal information of the user.
Biometric Data	Biometric data can be used for diagnosis activities in participatory sensing applications. Biometric data normally includes a participant's current physiological state and personal information, such as age and gender. Therefore, privacy will be leaked if the biometric data is identified.

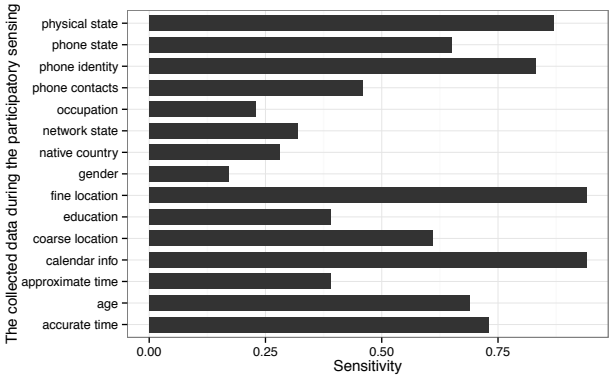


Fig. 3.4: The sensitivities of one participant to the set of data types.

coarse location information with around 0.35, whereas participant no. 28 has the highest sensitivity with almost 0.9. This rather large spread confirms our belief that privacy is fluid and its perception can strongly differ from person to person, making personalized privacy measurement approaches necessary.

3.6.2 Accuracy

Next, we evaluate the accuracy of the PriMe approach. For this, we compare the privacy measurement results generated by our approach with the participants' sensitivity statements (we discussed in the previous section). The similar the two results are, the higher the accuracy of PriMe is. To quantify the accuracy, we apply the Normalized Distance-based

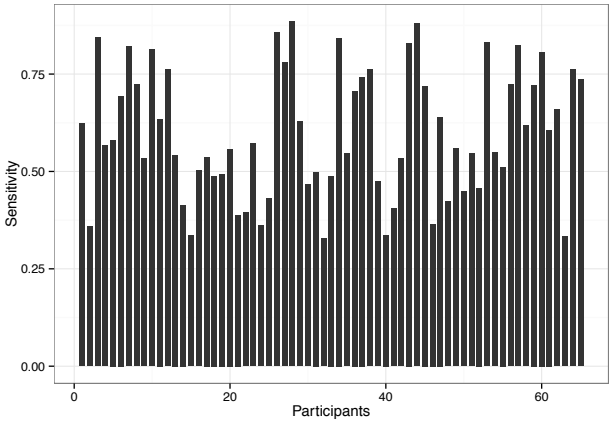
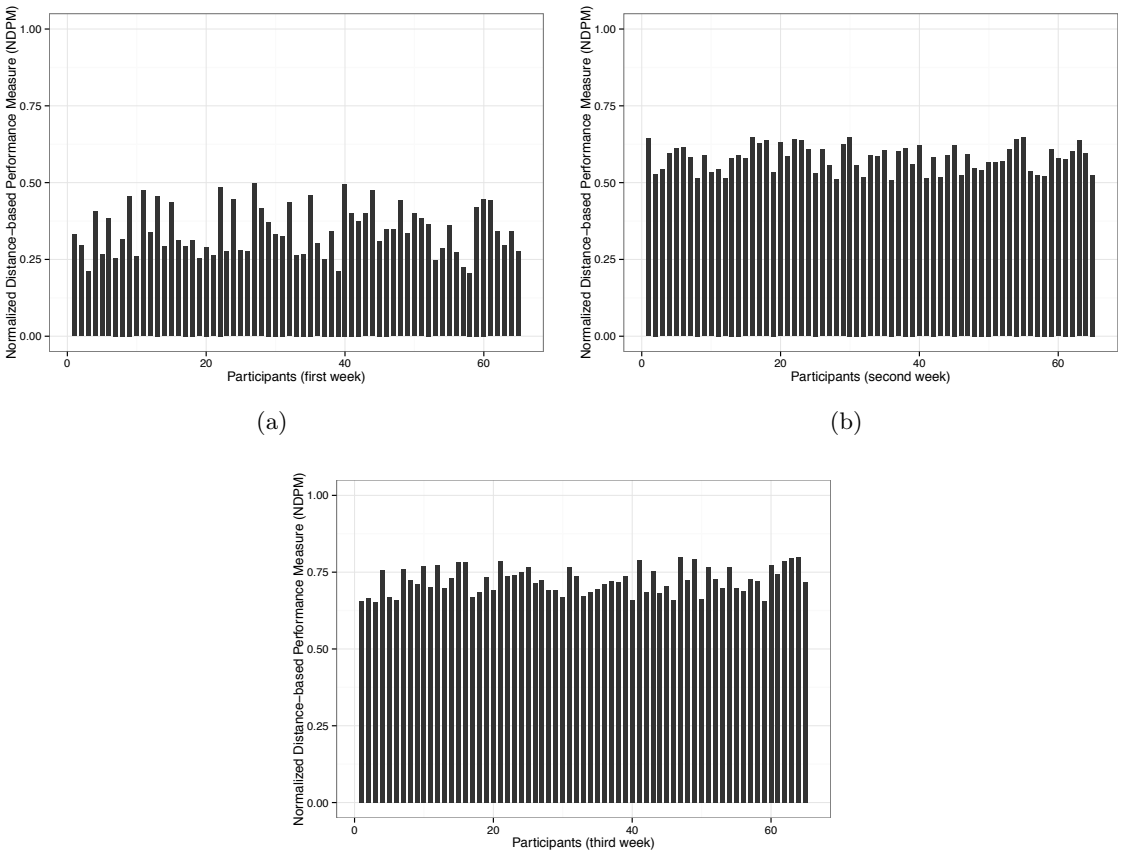


Fig. 3.5: The participants' sensitivity towards sharing coarse location data.



(a) (b)

(c)

Fig. 3.6: The accuracy of PriMe's results compared to the participants' statements using NDPM accumulated by study week.

Performance Measure (NDPM) approach [Yao95].

Fig 3.6 shows the results of the NDPM analysis, from week one to week three. The plots show that PriMe's accuracy increased over the duration of the study, achieving a good accuracy after the third week. Even in the worst case, PriMe predicts more than 60% of the participants' rankings. More specifically, the average accuracy is low in the first week due to the limited amount of feedback in the early phase of the study. However, the measurement results steadily increase as we obtain more user feedbacks with regard to their preferences over time. In the third week, we found the accuracy became good to high. In future work, we plan on conducting a longer study in order to further examine the accuracy improvement.

3.6.3 Trustworthiness

To test whether the participants trusted PriMe's assessment, we added a proxy function to the PriMe App that allows it to accept sensing tasks automatically on behalf of the users. This proxy function can be activated or deactivated at any time, which at least implicitly indicates the level of trust in the system. Fig. 3.7 shows the results of the proxy activation recordings. More than 50% of the participants activated the proxy function in the second week. This means they trust the results generated by PriMe after using it for a while. Further, more participants enabled the proxy function in the third week than disabled it. Approximately 18% of the participants did not use the function during the study.

Finally, after the study, we asked the participants to answer a questionnaire on how they felt with regard to their privacy in participatory sensing. Twenty-one participants responded. Many participants noted that PriMe App's explicit listing of which data is needed in order to fulfill a sensing task increased their awareness of their privacy concerns. As examples, we would like to share the following two characteristic comments from participants:

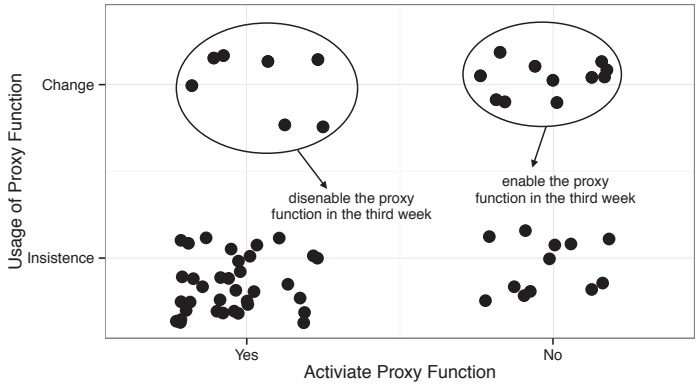


Fig. 3.7: Scatter plot showing the distribution of participants using the proxy function.

"At the beginning, I didn't care about my privacy at all when I accepted sensing tasks, but when I saw the data in my screen, I realized that some sensitive information may be disclosed."

"The participatory sensing App looks interesting and I also wanted to publish some tasks using it, but the big privacy risks really discouraged me."

3.7 Summary

In this chapter, we presented PriMe, a personalized privacy measurement method for mobile participatory sensing systems. Based on the proposed properties of privacy in participatory sensing, we measure the privacy from the perspective of an individual's attitude, which is represented by two intuitive properties, namely the intrinsic sensitivity, i.e., the individual inherent sensitivity, and the extrinsic sensitivity, i.e., the individual sensitivity to different data in different scenarios. The real world study with 65 users shows that PriMe provides reasonable and accurate results, and that the participants, in turn, trust the system to a high degree.

Although we have conducted a system to measure privacy of each user in mobile participatory sensing systems and tested it based on a user study, we acknowledge that it is

not the final step for this research. The ultimate objective is to help users to accept or reject tasks automatically based on their privacy concerns. To achieve this, we will refine our system and conduct more large-scale real-world tests (e.g., by releasing our App in App Stores) to get more reliable results. We will also try to provide more options, not only yes or no, to users to collect their preferences, which could increase the accuracy of PriMe with regard to the users' more fine-grained perception of privacy.

Chapter 4

Mitigating Privacy Risks of Mobile Apps using Crowdsourcing

In this chapter, we focus on mitigating privacy risks of smartphone applications using crowdsourcing-based recommendation approach. We propose and implement PriWe, a system based on crowdsourcing driven by users who contribute data access permission settings of their apps in the Android smartphone.

This chapter is organized as follows: Section 4.1 is the overview of this work. Users' expectation of smartphone privacy is illustrated in Section 4.2. We present our recommendation algorithm design in Section 4.3 and elaborate the design and implementation of the system in Section 4.4. Subsequently, to buttress our idea and system, we deploy PriWe in the real world and evaluation results are demonstrated in Section 4.5. We make a discussion in Section 4.6 and conclude this chapter in Section 4.7.

4.1 Overview

Mobile devices like smartphones or tablets are so popular today that billions of users all around the world are relying on them to handle personal and business affairs, such as emails, calendar management, entertainment, etc. Unfortunately, the widely adoption of such devices are coming with some potential privacy threats, as they have gained access to

lots of personal and sensitive data, such as user locations, contacts, and so on.

To mitigate such threats, system vendors have provided several mechanisms to confine the sensitive information accessible to mobile apps. For example, iOS from Apple has menu entries that enables users to control apps' permissions to sensitive data resources. For Android, one of the most popular mobile platform, its latest version (i.e., Android M released in May 2015) supports similar fine grained control on permission to replace its previous ineffective "all-or-none" scheme [and].

However, such fine grained control framework has its own drawbacks. For example, not all users have enough background knowledge to make the privacy configuration correctly. Also, there are so many apps and different permissions, so it is really a tedious job to require users to set all of them up. Finally, users hold different attitude to the privacy. They are willing to provide some information for better services and experiences, and meanwhile they are reluctant to share sensitive data due to privacy concern. To overcome the drawbacks, it is significant and beneficial to understand users' expectations of privacy and help them to set the privacy permission accordingly.

In this chapter, we propose a novel method that can help users finish their privacy settings properly and quickly. Our method is based on some key insights on how users decide whether to grant a permission to an app or not. First, the decision depends on a user's specific privacy preference or concerns, for example, whether a user cares more on geographical location than contact lists. Second, the decision is also related to a user's expectations on certain apps, for example, a user would expect an alarm app to access calendar, but would not expect that app to access his/her current geographical location. More details and discussion will be given in Section 4.2.

The method proposed in this chapter is first to learn the similarities among users in terms of privacy preferences and privacy expectations on apps, and then to recommend

appropriate permission settings to users based on such similarities. The rationale behind our method is that: users who share similar preferences on certain private data and/or privacy expectations on apps are more likely to make similar decisions in related privacy items.

To prove our proposed method, we have designed and implemented a system called *PriWe*, and evaluated it with lots of real world users (with 382 participants from Amazon Mechanical Turk, and 78 recruited volunteers from different cities). The results show that PriWe indeed is able to make proper recommendations that match users' privacy expectations and thus are mostly accepted by users.

Our Contributions. In this work we make following contributions:

- We proposed PriWe to understand users' expectation of privacy on mobile apps using the crowdsourcing mechanism.

- We proposed a novel recommendation approach, combining the item-based and user-based collaborative filtering methods for generating the recommendations for users' privacy permission settings.

- We implemented and deployed PriWe in the real world for evaluation. We collected the feedbacks of 78 users from the real world and 382 participants from the Amazon Mechanical Turk. According to the results, PriWe can make recommendations which are mostly accepted by users, thereby help them to make informed decisions and mitigate privacy disclosure.

4.2 Users' Expectation of Privacy

Taking a step back, we discuss the privacy in this section and figure out why understanding the individual expectation of privacy towards mobile apps is vital and beneficial.

In 1890, two U.S. lawyers proposed a prevalent definition, private life, habits, act, relations and the right to be alone [WB90]. With the proliferation of information technology, Wesin proposed that privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others, and it came to be known as information privacy [Wes68]. These two acknowledged definitions both emphasized that privacy to people should be an ability to express themselves selectively. The expression is driven by the individual expectation of privacy.

Another conceptual framework for understanding privacy expectations, called contextual integrity, suggests that privacy comprises appropriateness and distribution [Nis04]. More specifically, appropriateness focuses on whether the revelation of a particular piece of information is appropriate in a given context. For example, users' location data can be proffered in a map apps rather than in a game app. While distribution defines the occurrence of an information transfer from one party to another. For instance, one person is willing to share his/her data with friends instead of strangers. Therefore, there is a trade-off between services and privacy. People's expectation of privacy is just a reflection of such trade-off.

Using mobile apps is a typical scenario due to the discussion. Mobile services, especially smartphones, have become an important platform which can provide the multifarious services, integrating into our lives. There is almost no way to 100% protect users' information when they are using smartphones. More important, users also have ambivalent attitude to the data usage of mobile apps. They want to provide their data selectively based on their expectations to the privacy of mobile apps. On one hand, we yearn for better services and performances so that we are willing to provide some information. On the other hand, in general we are reluctant to share information because we also hope that our sensitive data could be preserved. Thus, understanding the users' privacy expectation on mobile apps is a key point for addressing the privacy issues.

It is a significant difference between our work and the existing research from security perspective [IAKR15]. More specifically, they assume there are the correct options for the privacy permission settings or other security issues. However, there is no absolute right answer in our design. All the privacy permission settings are based on the users' expectations. That is also the underpinning of our method and system.

4.3 Recommendation Mechanism

In this section, we elaborate the proposed recommendation algorithm, which takes advantage of users' demographic information and permission classification to improve the performance. The basic idea is presented in Section 4.3.1, followed by the item- and user-based collaborative filtering recommendation approaches in Section 4.3.2. Finally, we show how to conduct a hybrid recommendation in Section 4.3.3.

4.3.1 Overview

When you want to set the privacy permissions of each mobile app on your smartphone but you do not know how to set them appropriately, what will you do? Asking some other people or search Internet for suggestions may be an immediate and intuitive idea. And this is exactly what we proposed to do: our system will, on your behalf, go and collected opinions from a group of people who share similar backgrounds, privacy concerns and expectations, etc., and make the most appropriate recommendations to you.

A comprehensive investigation about recommendation system has been conducted [JZFF10]. According to the advantages of different recommendation algorithms and the characters of our scenario, we choose collaborative filtering methods to implement our idea. However, the recommendation mechanism is originally design to attract customers to buy commodities in some e-commerce markets, such as Amazon and Taobao. In our case, we do not have customers and commodities; rather we have smartphone users and privacy permission

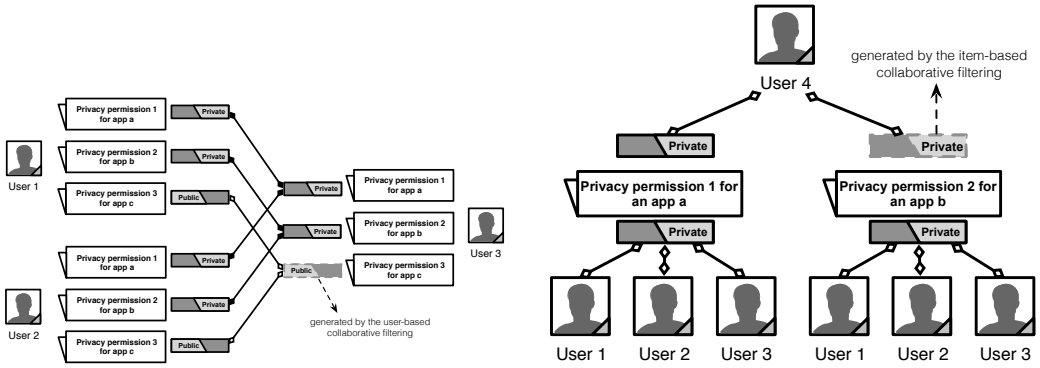
settings. We consider that people with similar backgrounds and habits may have similar privacy preferences. Thus, we map each smartphone user to a customer and each privacy permission setting to a commodity so that the item- and user-based collaborative filtering algorithms can play an expected role in our work. Further, we combine these two algorithms based on conditional probability with considering demographic and permission group information. Such hybrid algorithm can overcome the intrinsic drawbacks and achieve better performance of item- and user-based collaborative filtering algorithms

According to our discussion of privacy and the idea of the work, we initialized our recommendation algorithm through crowdsourced users' privacy permission settings rather than some experts' opinions. That is because we believe users' expectation should be the key to set the privacy permissions of their mobile apps.

4.3.2 Collaborative filtering

We assume that there are K users and each user has M apps. Each app holds N data access permissions. $r_{i,a,g}$ is defined as the setting of data permission g of the app a set by the user i. Users are allowed to set the privacy setting by the dichotomous variable {0, 1}. More specifically, $r_{i,a,g} = 0$ denotes that the users are averse to share the data with anyone, whereas $r_{i,a,g} = 1$ means the participant allows the disclosure of that information. However, the users may have sufficient understanding to different privacy permissions when they want to make a setting. It is also arduous for them to finish all of the privacy settings. To address this issue, we take advantage of user-based and item-based collaborative filtering algorithms. The following two examples and Fig. 4.1 further illustrate these two algorithms.

Example 1: Two users, i and j, both installed two apps a, b in the smartphone, and each app holds two permissions g, h. User i and j both allow app a to get the corresponding data permissions, by setting $r_{i,a,g} = 1$ & $r_{i,a,h} = 1$ and $r_{j,a,g} = 1$ & $r_{j,a,h} = 1$. In this situation, we consider they may have the similar privacy preferences. If user i set $r_{i,b,g} = 0$ to prohibit



(a) User-based collaborative filtering algorithm (b) Item-based collaborative filtering algorithm

Fig. 4.1: Generating recommendation of data access permissions for Android apps is based on the user- and item-based collaborative filtering algorithm.

the access permission g of app b, user j is likely to have the same choice on this setting.

Example 2: Two apps, a' and b', both are installed in the smartphone carried by user i' and user j'. The apps a' and b' hold the permissions g' and h', respectively. If users i' and j' both reject the data access, namely setting $r_{i',a',g'} = 0$ & $r_{i',b',h'} = 0$ and $r_{j',a',g'} = 0$ & $r_{j',b',h'} = 0$. In this case, permission g' of app a' and permission h' of app b' can be considered as two similar ones because they are both rejected by users i' and j'. The more users do this, the higher similarity the two permissions have. Thus, when newcomers have negative opinion to the privacy permission g' of app a', we also recommend them to reject the data access of permission h' of app $'b'$.

The example 1 and 2 illustrate the basic idea of user-based and item-based collaborative filtering approaches, respectively. According to the examples, finding similar privacy permissions and the users who have the similar privacy preferences. Thus, we show how to calculate the similarity of users and permissions. $s_u(i, j)$ is defined as the similarity between user i and user j. The similarity reflects how similar the users i and j are, i.e., how many privacy settings the two users have the same choice. The more such settings, the higher similarity between them. Thus, $s_u(i, j)$ is calculated through Eq. 4.1, based on

the Pearson correlation coefficient. The possible similarity values are between -1 and $+1$, where values near 1 indicate a strong similarity. We select Pearson correlation coefficient since the empirical analysis showed that for user-based recommender systems by far, the Pearson correlation coefficient outperforms other measures [HKBR99].

$$s_u(i,j) = \frac{\sum_{a\in M}\sum_{g\in N}(r_{i,a,g} - \overline{r}_i)(r_{j,a,g} - \overline{r}_j)}{\sqrt{\sum_{a\in M}\sum_{g\in N}(r_{i,a,g} - \bar{r_i})^2}\sqrt{\sum_{a\in M}\sum_{g\in N}(r_{j,a,g} - \bar{r_j})^2}} \tag{4.1}$$

We obtain the set of similar users by applying a threshold using $top - Q$ strategy. The $top - Q$ set of similar users to user i, $S_u(i)$ can be generated according to Eq. 4.2

$$S_u(i) = \{j | rank\ s_u(i,j) \leq Q\} \tag{4.2}$$

Likewise, we define $s_i(g,h)$ as the similarity between the privacy permission g and h. The similarity is based on the existing users' settings as illustrated in the Example 2. To calculate the similarity, we adopt the adjusted cosine similarity to take the differences of the average setting behaviors of the users into account, as shown in Eq. 4.3. We also select $top - Q$ similar items according to Eq. 4.4.

$$s_i(g,h) = \frac{\sum_{i\in K}\sum_{a\in M}(r_{i,a,g} - \overline{r}_i)(r_{i,a,h} - \overline{r}_i)}{\sqrt{\sum_{i\in K}\sum_{a\in M}(r_{i,a,g} - \overline{r}_i)^2}\sqrt{\sum_{i\in K}\sum_{a\in M}(r_{i,a,h} - \overline{r}_i)^2}} \tag{4.3}$$

$$S_i(g) = \{h | rank\ s_i(g,h) \leq Q\} \tag{4.4}$$

We have $\bar{r_i}$ now. it is the average permission setting for user i. The results for the adjusted cosine measure correspondingly range from -1 to $+1$. We adopt the adjusted cosine similarity to calculate the similarity between permission settings because it has been presented that the adjusted cosine similarity consistently outperforms the other metric in the item-based collaborative filtering approaches [HKBR99].

4.3.3 Fusion based on demographic and permission information

We propose a method that labels different users and permissions and then fuses the user- and item-based algorithms based on the labels. It is based on our observation that demographic and permission classification data can provide additional information about one specific user, thus can lead to better results in calculating similarities. Also, the demographic information and classification are getting more widely used in recommendation system, these information could also be deployed to fuse the user- and item-based collaborative filtering.

Thus, considering users, items, and labels jointly, we have a three-dimensional relation $< user, item, label >$. This three-dimensionalities can be projected as three two-dimensional relation, $< user, item >$, $< user, label >$, and $< item, label >$. In our case, we assume a set of user labels L_k and a set of permission labels L_n. More specifically, the user labels are generated according to demographic information, such as age, gender, occupation, and activity of mobile apps; item labels are based on the classification of permissions. Then, as shown in Fig. 4.2 the new set of users can be extend by item labels, $K' = K + L_n$ and the new set of items are extended by user labels, $N' = N + L_k$. Thus, the new matrix for recomputing the similarity $s_u(i, j)$ using user-based collaborative filtering is represented in a $K \times N'$ matrix, and the new matrix for recalculating the $s_i(g, h)$ using item-based collaborative filtering is denoted by a $K' \times N$ matrix.

We fuse the similarities $s_u(i, j)$ and $s_i(g, h)$ based on probability to generate a more robust similarity and overcome the data sparsity problem, which is an obstacle to our work in real-world deployment. More specifically, we provide different weights to the two similarities $s_u(i, j)$ and $s_i(g, h)$ and form a unified similarity. In this case, the user-based and item-based collaborative filtering approaches are only two special cases in the unified form.

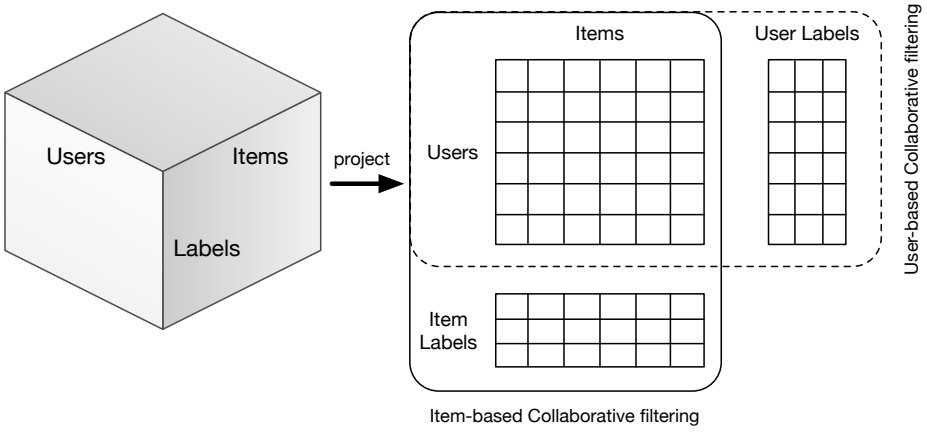


Fig. 4.2: The three-dimensional matrix *user-item-label* is projected as three two-dimensional matrixes, *user-item*, *user-label*, and *item-label*.

Assume we want to make a recommendation for user x about the privacy setting of permission z of app y, namely calculating $r_{x,y,z}$. In the light of previous illustration, user-based collaborative filtering approach only considers privacy settings provided by the users who have the similar privacy preference. Thus, we define the existing privacy settings for calculating $r_{x,y,z}$ as a set, US, $US_{x,y,z} = \{r_{i,y,z} | i \in S_u(x)\}$. Likewise, item-based collaborative filtering approach considers privacy settings to items. We also define a set, IS, $IS_{x,y,z} = \{r_{x,a,g} | g \in S_i(z)\}$. The similarity fusion algorithm considers these two sets jointly, i.e., UIS, $UIS_{x,y,z} = \{r_{i,a,g} | i \in S_u(x), g \in S_i(z)\}$.

When we scrutinize the privacy settings provided by users, we find they have totally different preferences to the same permission. Some users always have the sensitive attitude to the data usage, while others rarely shut down the permission due to their intrinsic traits. Some permissions are always shut down, simply because they have been set by some sensitive users. To eliminate such effect, we normalize the collected privacy settings by removing the average value, as shown in Eq. 4.5.

$$p_{x,y,z}(r_{i,a,g}) = r_{i,a,g} - (\overline{r}_i - \overline{r}_x) - (\overline{r}_{a,g} - \overline{r}_{y,z}) \tag{4.5}$$

$p_{x,y,z}(r_{i,a,g})$ serves as a normalizing function of the privacy setting of the permission z of the app y set by the user x, based on the existing crowdsourced privacy setting $r_{i,a,g}$. $\overline{r}_{a,g}$ and $\overline{r}_{y,z}$ are the mean of the privacy setting of permission g of app a and the privacy setting of permission z of app y, respectively. The sample space of the privacy permission settings should be defined as $\Phi_r = \{\emptyset, 0, 1, 2, ..., r\}$. In our case, there are actually three options, i.e., $\Phi_r = \{\emptyset, 0, 1\}$. $\emptyset$ means that the privacy settings have not been set so far, 0 expresses that users regard the information as private, and 1 presents that users allow the disclosure of this information. Therefore, $r_{i,a,g}$ denotes a privacy setting of permission g of app a, which is provided by user i, over the sample space Φ_r. Then, given a set of normalized settings, $\Omega_{x,y,z}$, we will be able to calculate the probability of $r_{x,y,z}$ by conditional on $\mathbb{P}(r_{x,y,z}|\Omega_{x,y,z})$, where Ω is given in Eq. 4.6.

$$\Omega_{x,y,z} = \{p_{x,y,z}(r_{i,a,g})|r_{i,a,g} \neq \emptyset\} \tag{4.6}$$

Now taking both user- and item-based recommendation algorithms into consideration, i.e $r_{i,a,g} \in (US, IS)$, we get the conditional probability presented in Eq. 4.7. That is, if we know $r_{i,a,g} \in (US, IS)$, we can eventually obtain the conditional probability of $r_{x,y,z}$, conditioning on the set Ω.

$$\mathbb{P}(r_{x,y,z}|\Omega_{x,y,z}) = \mathbb{P}(r_{x,y,z}|\{p_{x,y,z}(r_{i,a,g})|r_{i,a,g} \in US \cup IS\}) \tag{4.7}$$

Eq. 4.7 indicates that the probability of $r_{x,y,z}$ depends only on $r_{i,a,g}$. Thus we can write Eq. 4.7 for short as $P(r_{x,y,z}|\Omega_{x,y,z}) = P(r_{x,y,z}|r_{i,a,g} \in US \cup IS)$. We introduce two independent binary indicators I_1 and I_2 to present the dependency of $r_{i,a,g}$ on set US and IS. That is, $I_1 = 1$ corresponds to dependency on the set US while $I_1 = 0$ indicates independency. Likewise, $I_2 = 1$ states $r_{i,a,g}$ depends on the set IS while $I_2 = 0$ indicates $r_{i,a,g}$ is independent of IS. Therefore, given the two sets US and IS, we can derive Eq. 4.8

based on the indicator I_1 and I_2.

$$
\begin{aligned}
&\mathbb{P}(r_{x,y,z}|US, IS) \\
&= \sum_{I_1}\sum_{I_2} \mathbb{P}(r_{x,y,z}|I_1, I_2, US, IS)\mathbb{P}(I_1, I_2|US, IS) \\
&= \mathbb{P}(r_{x,y,z}|I_1 = 0, I_2 = 0, US, IS)\mathbb{P}(I_1 = 0, I_2 = 0|US, IS) \\
&+ \mathbb{P}(r_{x,y,z}|I_1 = 1, I_2 = 0, US, IS)\mathbb{P}(I_1 = 1, I_2 = 0|US, IS) \\
&+ \mathbb{P}(r_{x,y,z}|I_1 = 0, I_2 = 1, US, IS)\mathbb{P}(I_1 = 0, I_2 = 1|US, IS) \\
&+ \mathbb{P}(r_{x,y,z}|I_1 = 1, I_2 = 1, US, IS)\mathbb{P}(I_1 = 1, I_2 = 1|US, IS)
\end{aligned} \tag{4.8}
$$

Based on the definition of indicators I_1, I_2, $r_{i,a,g}$ is independent to US if $I_1 = 0$ and is irrelevant to IS when $I_2 = 0$. Thus, $\mathbb{P}(r_{x,y,z}|I_1 = 1, I_2 = 0, US, IS) = \mathbb{P}(r_{x,y,z}|US)$, and $\mathbb{P}(r_{x,y,z}|I_1 = 0, I_2 = 1, US, IS) = \mathbb{P}(r_{x,y,z}|IS)$. Obviously, we cannot generate any recommendation without the sets US and IS, which means $\mathbb{P}(r_{x,y,z}|I_1 = 0, I_2 = 0, US, IS) = 0$. When we consider the sets US and IS jointly, this two sets can be regarded as a new set UIS. Namely, $\mathbb{P}(r_{x,y,z}|I_1 = 1, I_2 = 1, US, IS) = \mathbb{P}(r_{x,y,z}|UIS)$. Thus, we can obtain Eq. 4.9.

$$
\begin{aligned}
\mathbb{P}(r_{x,y,z}|US, IS) &= \mathbb{P}(r_{x,y,z}|US)\mathbb{P}(I_1 = 1, I_2 = 0|US, IS) \\
&+ \mathbb{P}(r_{x,y,z}|IS)\mathbb{P}(I_1 = 0, I_2 = 1|US, IS) \\
&+ \mathbb{P}(r_{x,y,z}|UIS)\mathbb{P}(I_1 = 1, I_2 = 1|US, IS)
\end{aligned} \tag{4.9}
$$

For easy computation, we use two parameters λ and δ in Eq. 4.10, assuming $\mathbb{P}(I_1 = 1|US, IS) = \lambda$ and $\mathbb{P}(I_2 = 1|US, IS) = \delta$. According to Eq. 4.10, the $r_{i,a,g}$ depends on both sets US and IS, i.e., UIS, when $\lambda = 1$ and $\delta = 1$. Likewise, the $r_{i,a,g}$ has 0.5 probability

dependent on US, if $\lambda = 0.5$; the set IS also can play a half role when δ is 0.5.

$$\begin{aligned}\mathbb{P}(r_{x,y,z}|US, IS) = \mathbb{P}(r_{x,y,z}|US)\lambda(1-\delta) \\ + \mathbb{P}(r_{x,y,z}|IS)(1-\lambda)\delta \\ + \mathbb{P}(r_{x,y,z}|UIS)\lambda\delta \end{aligned} \tag{4.10}$$

Afterwards, we can get the estimated privacy settings $r_{x,y,z}$, as presented in Eq. 4.11. We can determine the parameters λ and δ through iterations in the experiments.

$$\begin{aligned}\widehat{r}_{x,y,z} &= \sum_{t=1}^{\Phi_r} t\mathbb{P}(r_{x,y,z} = t|US, IS) \\ &= \big(\sum_{t=1}^{\Phi_r} t\mathbb{P}(r_{x,y,z} = t|UIS)\lambda\delta\big) \\ &+ \big(\sum_{t=1}^{\Phi_r} t\mathbb{P}(r_{x,y,z} = t|US)\lambda(1-\delta)\big) \\ &+ \big(\sum_{t=1}^{\Phi_r} t\mathbb{P}(r_{x,y,z} = t|IS)(1-\lambda)\delta\big)\end{aligned} \tag{4.11}$$

Now we need to estimate the conditional probability in Eq. 4.11, namely, $\mathbb{P}(r_{x,y,z} = t|UIS)$, $\mathbb{P}(r_{x,y,z} = t|US)$, and $\mathbb{P}(r_{x,y,z} = t|IS)$. The basic idea of the estimation is to calculate the likelihood of $r_{x,y,z}$ to be similar with $r_{i,a,g}$ based on the sets US, UI, and UIS. Hence, we make use of the similarity between users to calculate the likelihood based on US, as shown in Eq. 4.12. Likewise, the similarity function $s_i(.)$ is used to compute the likelihood based on the set IS, as presented in Eq. 4.13.

$$\mathbb{P}(r_{x,y,z} = t|US) = \frac{\sum_{\forall r_{i,a,g}:(r_{i,a,g}\in US)\wedge(r_{x,y,z}=t)} s_u(i,x)}{\sum_{\forall r_{i,a,g}:r_{i,a,g}\in US} s_u(i,x)} \tag{4.12}$$

$$\mathbb{P}(r_{x,y,z} = t|IS) = \frac{\sum_{\forall r_{i,u,a}:(r_{i,a,g}\in IS)\wedge(r_{x,y,z}=t)} s_i(g,z)}{\sum_{\forall r_{i,a,g}:r_{i,a,g}\in IS} s_i(g,z)} \tag{4.13}$$

Calculating the likelihood based on UIS is a little tricky. We consider the probability estimation as the combination of the similarity function $s_u(.)$ and $s_i(.)$. More specifically, we use Euclidean distance to produce the similarity function, as illustrated in Eq. 4.15.

$$\begin{aligned}&\mathbb{P}(r_{x,y,z} = t|UIS) = \\ &= \frac{\sum_{\forall r_{i,a,g}:(r_{i,a,g}\in UIS)\wedge(r_{x,y,z}=t)} s_{ui}(r_{i,a,g}, r_{x,y,z})}{\sum_{\forall r_{i,a,g}:r_{i,a,g}\in UIS} s_{ui}(r_{i,a,g}, r_{x,y,z})}\end{aligned} \tag{4.14}$$

$$s_{ui}(r_{i,a,g}, r_{x,y,z}) = \frac{1}{\sqrt{(\frac{1}{s_u(i,x)})^2 + (\frac{1}{s_i(g,z)})^2}} \tag{4.15}$$

Now, we can get the results,

$$\widehat{r}_{x,y,z} = \sum_{r_{i,a,g}} p_{x,y,z}(r_{i,a,g}) W^{i,a,g}_{x,y,z} \tag{4.16}$$

where

$$W^{i,a,g}_{x,y,z} = \begin{cases} \frac{s_u(i,x)}{\sum_{r_{i,a,g}\in US} s_u(i,x)}\lambda(1-\delta) & r_{i,a,g} \in US \\ \frac{s_i(g,z)}{\sum_{r_{i,a,g}\in IS} s_i(g,z)}(1-\lambda)\delta & r_{i,a,g} \in IS \\ \frac{s_{ui}(r_{i,a,g},r_{x,y,z})}{\sum_{r_{i,a,g}\in UIS} s_{ui}(r_{i,a,g},r_{x,y,z})}\lambda\delta & r_{i,a,g} \in UIS \end{cases} \tag{4.17}$$

So far, we have elaborated the process of recommendation based on the crowdsourced privacy settings. The only thing is to determine the parameters λ and δ. When we deploy the system in the real world, we find these two parameters, λ and δ, reaching their optimal at 0.7 and 0.5, respectively. According to the illustration of the algorithm, the parameters are determined by the dataset, which means they are adaptive. More details are presented in Section 4.5.3.

4.4 System Design and Implementation

In this section, we illustrate the design of PriWe and the corresponding implementation.

4.4.1 Architecture

We have two intentions in our mind when designing PriWe. First, PriWe can help users to make better decisions on privacy settings in their own smartphones. Second, the processes of analyzing crowdsourced data and generating recommendations should be completed in a server due to the limited capability of smartphones. To achieve these intentions, we design the system, as illustrated in Fig. 4.3.

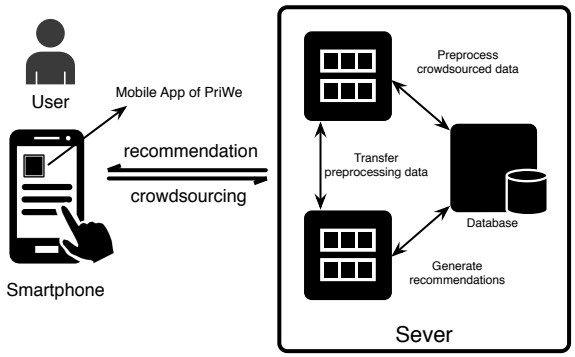


Fig. 4.3: The overview of PriWe, which insists of an mobile app in the smartphone and a server.

A mobile app is deployed in the smartphone to collect privacy settings from users. The mobile app of PriWe should consist of several features and provide various user-interfaces to interact with users. Firstly, it can automatically scan the apps installed in the smartphone and identify them by names. The user can browse the privacy permission settings of each app accordingly. Secondly, the PriWe app can apply the recommendations generated by the server. For example, the app allows users to set/change the privacy permission of each mobile app installed in the smartphones, it also can set the privacy permission automatically when users attempt to apply the recommendations provided the server. Finally, the PriWe app itself should hold the data access permissions as few as possible. Because the ultimate objective of our project is to help users to make better decisions for privacy settings in their smartphone and mitigate privacy risk accordingly, our system should be a privacy risk in

no event.

The server side of PriWe has two key components, which are responsible for preprocessing the crowdsourced data and generating the recommendations, respectively. More specifically, the former components aims to preprocess the collected data, such as validation and classification; the latter one mainly focuses on generating recommendations for various mobile apps of different users. The proposed recommendation algorithm is deployed in this component. All the information, including the raw crowdsourced data and processed results will be stored in an inbuilt database. The two components play a pivotal role in the server side and we elaborate on it subsequently.

According the design of mobile app and server side, we implement PriWe, as presented in Fig. 4.4.

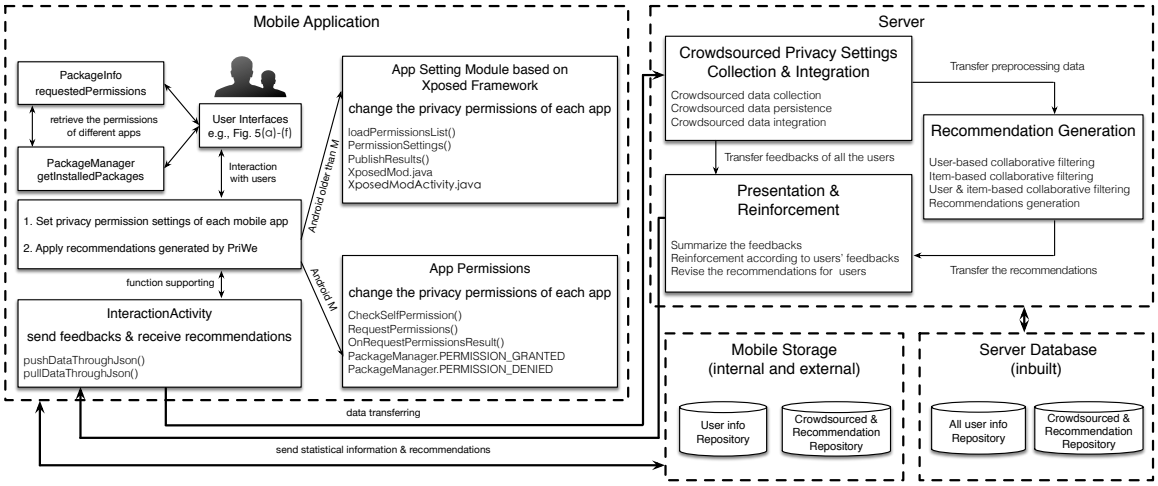


Fig. 4.4: The implementation architecture of PriWe

4.4.2 PriWe App

The mobile app of PriWe is developed in Android platform and can be compatible with various smartphones with different screen sizes. Our prototype of PriWe app is implemented on Android 4.4 and runs on the Google/LG Nexus 4 handset. For deployment in the real

world, PriWe has been tested with Android operating system 4.0.3 - 5.1.1.

There are two major objectives of the PriWe mobile app. The first is that it should provide various user-interface that can enable a user to set or change permission settings related to privacy, and at the same time, it should be able to send the settings back to its server. The second one is that the PriWe app is able to apply the recommendations generated by the server. It can set the privacy permissions automatically when a user has confirm to take the recommended settings. Furthermore, the PriWe app should also include a sign-in mechanism and be able to present the results. As shown in Fig. 4.4, users can browse what apps have been installed in their smartphones, and what data access permission the apps have held, through the APIs provided by official Android SDK. For example, function *PackageManager.getInstalledPackages(0)* can retrieve installed apps in the smartphone. Function *PackageInfo.requestedPermissions* can scrutinize the privacy permissions of each app. Such a method will return all data usages of the app. However, it is arduous for users to read all the system permissions in a screen of smartphone. we summarize eleven types of abused data and permissions of Android apps and discuss their potential risks, as shown in Table 4.1. The summary is based on some freeform comments on the Internet [mos13, top14], research papers on the Android system and analysis of smartphone apps [OTK+12, KCC+12, FHE+12, DMC14, FCH+11], security and privacy tips from official guidelines [tan14], a survey on information security and privacy of Android apps [JZ13].

Then, PriWe app allows them to set/change the permission settings or take the recommendations. If the user is using the Android system older than 5.0, PriWe app can trigger the App Setting Module, which is based on Xposed Framework [xpo] since such Android systems did not provide any mechanism for normal users to modify the privacy settings. If Android M is the current system, PriWe app will invoke the functions provided by the

Android M to change the data access permission of each app. These two modules allow users to change the privacy permission settings for various mobile apps. Since apps may be crashed when they cannot access specific data, PriWe feeds Android apps artificial data. However, there are two exceptions: access to the internet and modify external storage are restricted by denying access due to Android system mechanism. Overall, PriWe has the capability to modify the data access permission of installed apps in Android smartphone. Thus, PriWe requires root permission, namely, there is no way to achieve our objectives in non-root devices. Although root process is considered as legal, it is not supported officially. We take this issue in a neutral way and we do not advocate rooting Android smartphone for protecting users' privacy. However, in our work, PriWe needs root permission to mitigate information disclosure. Furthermore, according to the feedbacks from users, we did find no users have reported issues about data leakage or system crashed due to rooting their smartphones.

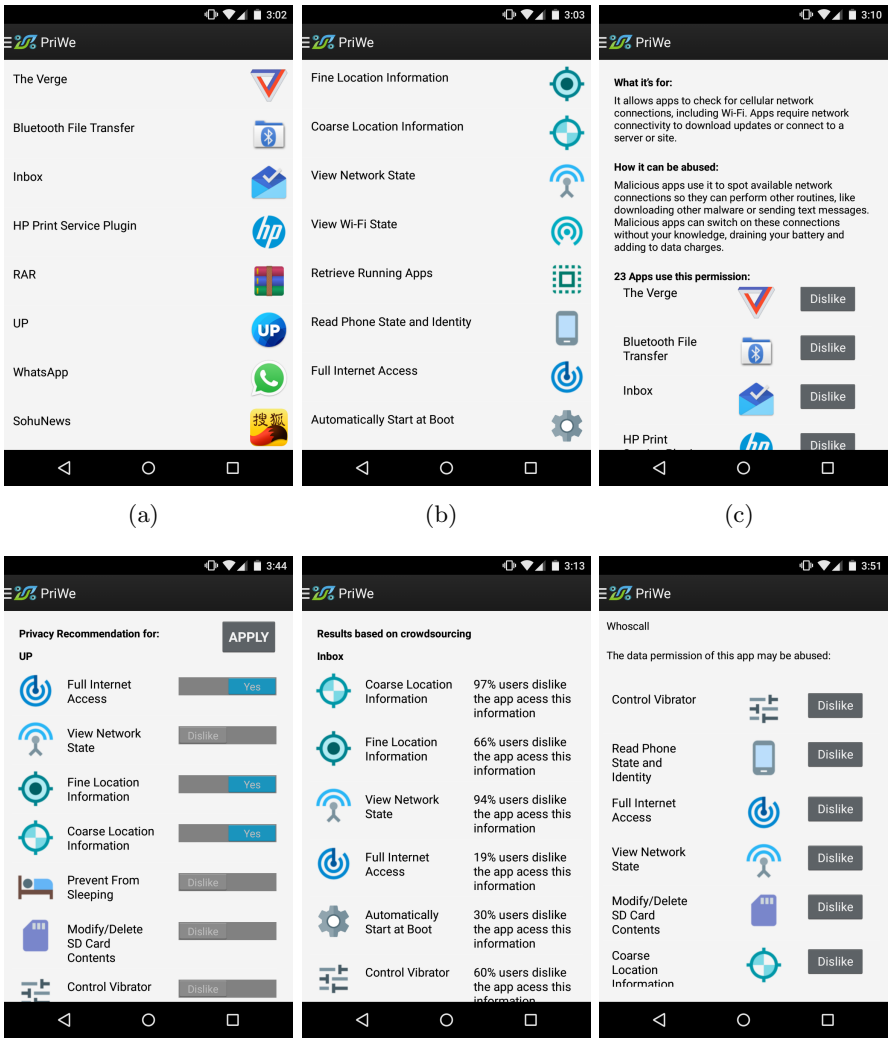
The key functions of PriWe mobile app are depicted in Fig. 4.5, which is composed by snapshots of the app in Nexus 4. All functions and interactions between them are implemented by Activity and Fragment, which are also provided by Android SDK.

4.4.3 PriWe Server

The server is designed to analyze collected data and generate recommendations according to crowdsourced data. As shown in Fig. 4.4, there are three key components in the server, which are responsible for data preprocessing, recommendation generation, and presentation and reinforcement. In the data preprocessing part, the server mainly focuses on cleaning and structuring the collected data, which will be the input of the next step, i.e., the recommendation generation part. To generate recommendations, the server applies the proposed method depicted in section **??**. The output of this step is a privacy permission setting list. In the presentation and reinforcement part, the server can summarize the collected data and

Table 4.1: Summary of most abused data and permissions

Most Abused Data and Permissions
• **Coarse and fine location** (Approximate or exact location information. It can lead location-based attacks or malware, or sending location-based ads.)
• **Network state** (Cellular network information and connections. It will also drain smartphones' battery.)
• **Wifi network information** (Wi-Fi network information, including passwords and usernames. It can lead information disclosure by Wi-Fi network.)
• **Running apps information** (Information of running tasks and processes. Users' sensitive information from other running apps can be leak.)
• **Phone state and identity** (Phone states information and International Mobile Equipment Identity. It can lead sensitive information disclosure.)
• **Modify/Delete internal/external contents** (Permission of modification internal and external storage. Apps steal information or save data on internal and external storage.)
• **Full internet access** (Permission of using the Internet to download and upload. The sensitive information can be disclosed and malware will be downloaded.)
• **Automatically Start at Boot** (Permission of automatically starting the smartphones boot. Malicious apps will use it to boot automatically.)
• **Send SMS Messages** (Permission of sending text messages without users' awareness for subscribe additional services which may leave users with unexpected charges.)
• **Prevent From Sleeping** (Permission of preventing from sleeping or the screen from dimming. Apps can steal the information even it is time-consuming.)
• **Control Vibrator** (Permission of accessing vibrator function. It can stop vibrations for notification before malicious apps interpret information.)



(a) (b) (c)

(d) (e) (f)

Fig. 4.5: PriWe provides an Android app for participants. (a) PriWe can scan various app installed in smartphones; (b) PriWe also provides an user interface to the participants to list the most abused data access permissions; (c) The participants can discover how many installed apps used a specific permission and provide their privacy preferences; (d) The participants can also take a look about how many permissions an app will use and show their feedbacks of privacy preference accordingly; (e) The statistical results are presented to the participants, which can be taken as a reference for their privacy preferences; and (f) PriWe can make recommendations to various apps according to the individual privacy preferences.

provide the corresponding statistics, which are also viewed by the users. The reinforcement part will revise the recommendation list if the user changed the permission settings even after accepting the recommendations.

The server system is deployed in an IBM server and built as three-tier architecture which is composed of an application tier, a domain logic tier, and a data persistence tier. More specifically, the application tier, is a web-front which implemented by Html, Javascript and the third development libraries. A user friendly interface can be provided in this tier. The domain logic tier is implemented by Java EE architecture and Enterprise Beans mechanism to analyze the collected data. To improve robustness and configurability of the system, the web application is built based on frameworks including Spring, Struts, Hibernate. The recommendation algorithm is also deployed in this tier to generate recommendations to the users. In data persistence tier, all data are persisted in MySql database.

4.5 Experiment and Evaluation

In this section, we demonstrate evaluation of PriWe, including experiment setup, data collection, results and our findings. We conducted two experiments to evaluate PriWe, one is based on the Amazon Mechanical Turk, the other one is based on the deployment in the real world. More specifically, we published a task to collect people's feedbacks about the privacy of smartphones on the Amazon Mechanical Turk. Furthermore, to make the evaluation results more convincing, we also deployed the PriWe in the real world. Subsequently, we elaborate on the evaluation from these two parts, respectively.

To evaluate the results of PriWe quantitatively, we proposed a metric to illustrate the accuracy of the recommendations generated by PriWe, as shown in Eq. 4.18. R_p denotes all the privacy permission settings the participants have chosen in the Amazon Mechanical Turk. R_i represents the recommendations of the corresponding privacy permission settings

provided by PriWe.

$$Accuracy(i) = \frac{R_p \cap R_i}{R_i} \tag{4.18}$$

4.5.1 Evaluation based on Amazon Mechanical Turk

We published a task on the Amazon Mechanical Turk[1] for three weeks, and 382 participants completed our task. In the task, we asked the participants to answer a questionnaire to illustrate their privacy preferences about various types of mobile apps. We prepared two questionnaires, survey A and survey B. Survey A investigated the privacy preferences of participants to various apps widely, while Survey B collected fine-grained participants' preferences to some particular mobile apps. 200 participants completed the survey A and 182 participants finished the survey B. To avoid bias and make the results more convincing, we present the statistics of the participants. Among all the participants, 243 participants are male, and 139 participants are female. 226 participants are 20-29 years old, and 115 participants are 30-39 years old. The remainder of the participants are either 10-19 or above 40. All of the participants came from various backgrounds, such as, energy, materials, consumer staples, health care, finance, information technology and etc. More information about the distribution of the participants in survey A and survey B can be seen from Table 4.2. Again, the distribution of the participants illustrates that we avoid the statistical bias and make the results convincing.

To evaluate the accuracy of recommendations produced by PriWe, we separate the survey A into two parts, one is regarded as a train set, the other one is regarded as a test set. So does the survey B. Furthermore, for more convincing and without bias, we also treated the survey A as a train set and the survey B as the test set and vice versa. All the results are demonstrated in Fig. 4.6. The overall accuracy of the recommendations made by PriWe is

[1] `https://www.mturk.com/mturk/preview?groupId=3PBTVBPQ8T1PENG33V3IMPSHIB9LG1`

Table 4.2: Statistics of participants in Amazon Mechanical Turk

Participants	Numbers in Survey A	Percentage in Survey A	Numbers in Survey B	Percentage in Survey B
Male	133	66.5%	110	60.4%
Female	67	33.5%	72	39.6%
10-19	4	2%	6	3.3%
20-24	45	22.5%	43	23.6%
25-29	69	34.5%	70	38.5%
30-40	64	32%	51	28%
40+	18	9%	12	6.6%
Energy	9	4.5%	6	3.3%
Materials	4	2%	6	3.3%
Industrials	19	9.5%	22	12.1%
Consumer Discretionary	13	6.5%	7	3.9%
Consumer Staples	12	6%	17	9.3%
Health Care	24	12%	17	9.3%
Finance	28	14%	21	11.5%
IT in Security & Privacy	27	13.5%	25	13.7%
IT in non Security & Privacy	40	20%	39	21.4%
Tele Services	15	7.5%	19	10.4%
Utilities	9	4.5%	3	1.7%
Rarely (0 1hr)	7	3.5%	9	4.9%
Sometimes (1 2hr)	49	24.5%	49	26.9%
Frequently (2 4 hr)	79	39.5%	56	30.8%
Very often (4+ hr)	65	32.5%	68	37.4%
Socializing	78	39%	59	32.4%
Shopping	23	11.5%	16	8.8%
Accomplishing	10	5%	14	7.7%
Arrangement	11	5.5%	13	7.1%
Discovery	25	12.5%	22	12.1%
Me Time	41	20.5%	35	19.2%
Self-expression	12	6%	23	12.6%
Deliberately completed	113	56.5%	119	65.4%
Normally completed	80	40%	61	33.5%
Hastily completed	7	3.5%	2	1.1%

about 78%. It indicates PriWe can make accurate and appropriate recommendations, which are mostly accepted. According to the results, the results based on survey A and survey B jointly are better than those based on either survey A or survey B. It indicates the recommendations can achieve higher accuracy when the data set consist of more crowdsourced permission settings. The combination of two surveys can also overcome the data sparsity issues in some degree.

We presented the results according to participants' gender, age, background, time spent on smartphones, favourite activity on smartphones and attitude to the survey, as shown in Fig. 4.6(a)-4.6(f). Fig. 4.6(a) demonstrates the recommendations provided by PriWe for male participants can achieve slightly higher accuracy than those for females. There is no obvious evidence to support that male have better understanding to the privacy permission of mobile apps. However, what we found is that when the participants are female, their most frequent activities on the smartphone are shopping and socializing. It may suggest that female users did not have enough attentions on the personal information on the smartphone. Another finding is that accuracy become high gradually with the increase of participants ages. One potential explanation is that some young people have no unambiguous perceptions about their privacy permission of their mobile apps. We investigate the participants whose background in information technology with a focus on privacy & security and other areas in information technology. The accuracy of recommendation for the participants in privacy and security is higher than the remainder of all the selected participants (around 90%), because the users who have the background about the information privacy and security have a better understanding about the privacy permission settings in smartphones. Due to the same reason, the users who came from other areas have lowest accuracy of recommendations. Fig. 4.6(d) indicates the PriWe did not provide so proper advices to the people who spent less time on the smartphone. They may have inadequate knowledge to the devices which

did not cost them much time. As shown in Fig. 4.6(e), people who like to use some accomplishing (e.g., managing finances, health and productivity) or arrangement (planning for upcoming events) apps will get more accurate recommendations from PriWe due to their existing and crowdsourced permission settings. In the last subfigure Fig. 4.6(f), we can see the people who completed our task in Amazon Mechanical Turk in a rush cannot get the accurate recommendations for their privacy permission settings since they just finish the task without any attention.

4.5.2 Evaluation based on real-world deployment

PriWe app has also been released to 78 users, who are from Hong Kong, Singapore, Austria, England, America and China, for evaluation in the real world. The server collected users' feedbacks of their permission settings and some basic information. The collected information includes app information, users' permission settings of installed apps, and the users' basic information, such as background, age, gender, user ID and etc. In the evaluation, we collected information from 78 participants based on PriWe. Since users have multifarious apps, the summary of number apps of each user is shown in Table 4.3. From the table, it can be seen that the majority of the users have less than 40 apps in their smartphone, which almost meet the statistic of users' apps from Statistics Portal [sta14].

Table 4.3: Statistics of participants' Android apps

Number of apps	Number of users	Percentage
1∼20	26	33%
20∼40	27	35%
40∼60	17	22%
60∼	8	10%

To corroborate the proposed abused data and permissions list, we calculate the average number of Android apps that participants installed access these data and permissions. According to the results as presented in Table 4.4, we found that all the potential abused

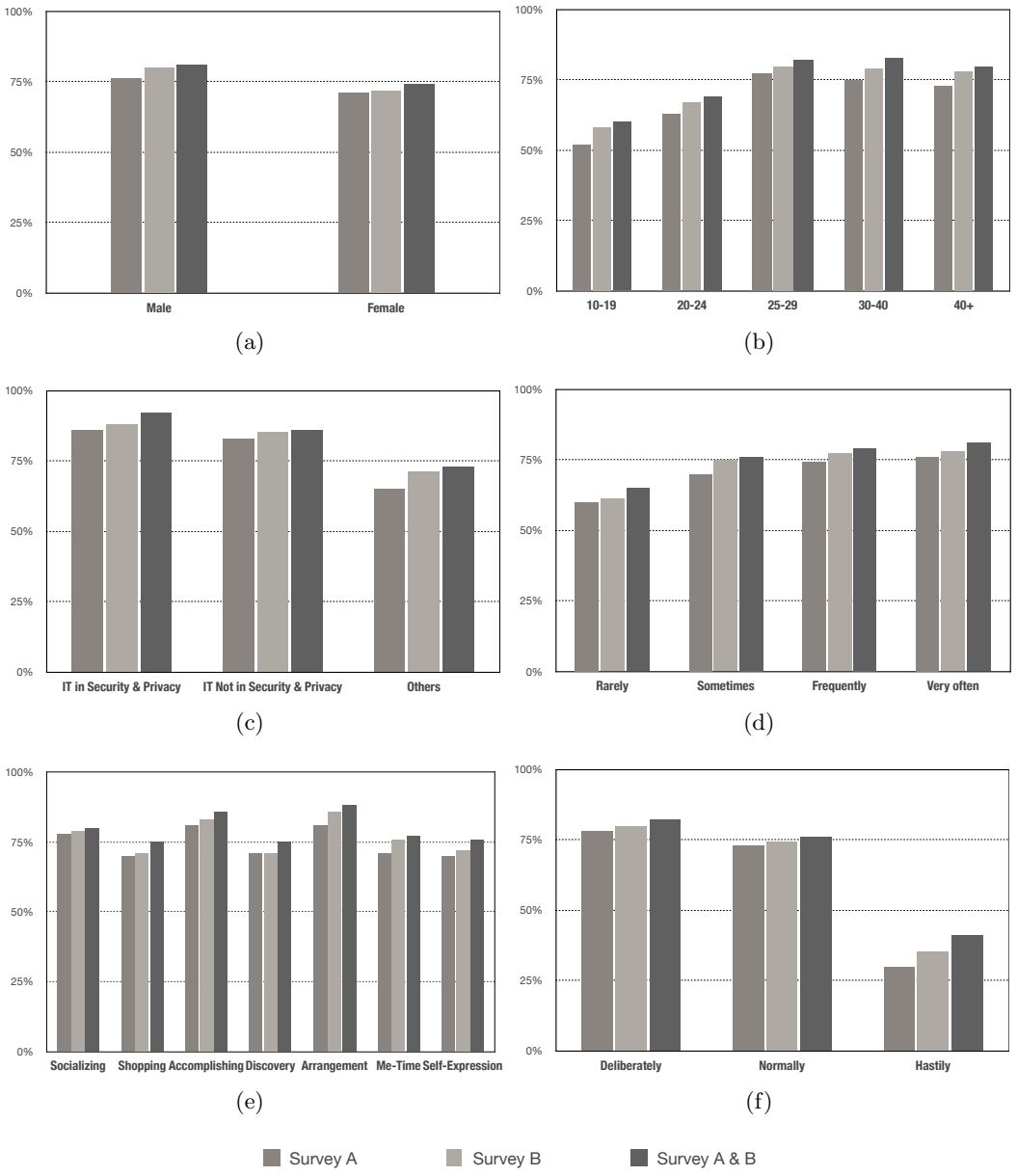


Fig. 4.6: The accuracy of recommendation generated by PriWe based on the participants' feedbacks in Amazon Mechanical Turk. The results are presented according to (a) the participants' genders (b) the participants' ages (c) the participants' backgrounds (d) the time participants spent on the smartphone (e) the most frequent activities of participants and (f) the attitudes of participants

privacy permissions have been accessed by many apps. These apps account large proportion of all the apps in the light of Table 4.3.

Table 4.4: The average number of Android apps that access abused inforamtion

Abused data and permissions	Number of apps
Coarse and fine location	16
Network state	32
Wifi network information	20
Running apps information	13
Phone state and identity	18
Modify/Delete contents	30
Full internet access	35
Automatically start at boot	17
Send SMS messages	7
Prevent from sleeping	25
Control vibrator	27
Access 2∼5	27
Access 6∼10	16
Access all	5

Since there is no clear or existing metric to evaluate our work, we treat the survey as the ground truth to evaluate PriWe. The results would be unconvincing if we did not take the participants' consideration.

We illustrated the evaluation results in Fig. 4.7 and Fig. 4.8 respectively. From Fig. 4.7, we can see that the recommendations are usually taken by the users. However, the recommendations about preventing from sleeping and controlling vibrator are not fully apprehended and reluctant to be applied by users. The reason of this phenomenon may be that they are not very severe risks and participants did not take much attention to them, ignoring the preferences and recommendations. The recommendations about location, network state and wifi network information, running apps and automatically starting are highly accepted. Participants may take them seriously since these information involved personal and even sensitive data. That is a reason why participants are willing to take them. Furthermore, participants showed ambivalence about the recommendations of phone

state and identity, modify storage contents, Internet capability and SMS Messages control. Since these information or permissions play important roles in apps running and service performances, the ambivalence presents participants hope to obtain better services and preserved these information as well.

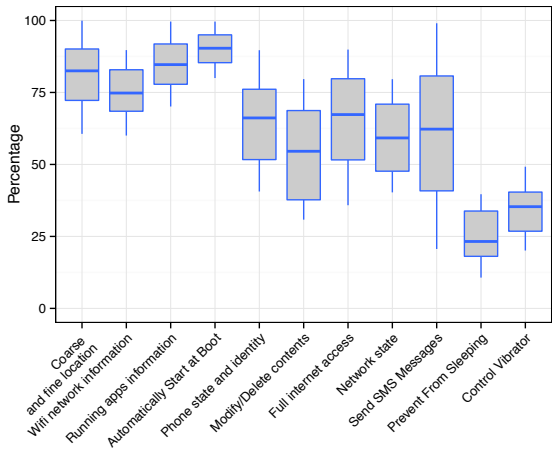


Fig. 4.7: The percentage of apps that users take the recommendations of each data permission.

To evaluate our another objective, i.e., improving awareness of privacy preference, we depicted the results according to the feedbacks in Fig. 4.8. From the graph, we can see that participants have a better comprehension or even epiphany to some privacy permissions. However, the participants did not have a better understanding about the permission of automatically boot and wifi network information. According to the survey after the experiment, we discovered that most participants already knew some mobile apps can boot automatically so they did not pay more attention to it. The wifi network is permeating our life in every aspects inevitably and people take it as a kind of routine. Thus, participants did not feel remarkable improvement of awareness of wifi network information.

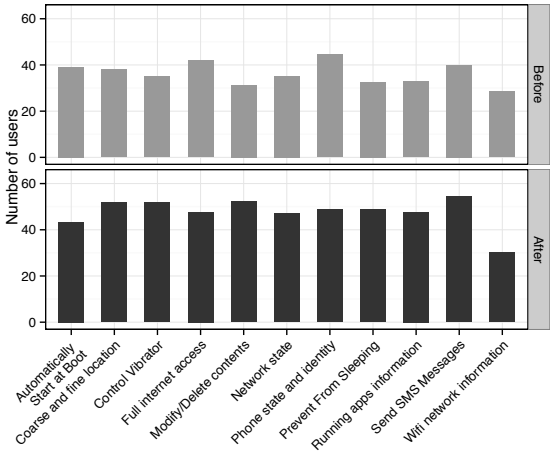


Fig. 4.8: The number of users have a better understanding of each data access permission after using PriWe.

4.5.3 Parameters estimation

There are two parameters, λ and δ, in the Eq. 4.17. Note that the two parameters are adaptive to deployment scenarios. We present the impact and estimation of the parameters in our case for better understanding. Furthermore, they are around 400 participants in the experiment and we make use of half of them to train the algorithm, namely the training dataset is 200. According to the convention [Ric06], it is adequate to build the algorithm and our discussion corroborate the premise as well.

We test the parameters according to the algorithm's performance. λ and δ represent the recommendations' dependence on the dataset US and IS, respectively. For consistency with experiments reported in the literatures [XLY+05, JZFF10], we take the mean absolute error (MAE) as the metric to evaluate the impact of two parameters, as shown in Equation 4.19. In the equation, L denotes the total number of predicted permission setting. The basic idea of MAE is to calculate the average absolute deviation of predictions to the ground truth data. In our case, we compute the deviation of our recommendation results to the actual

selections of the participants.

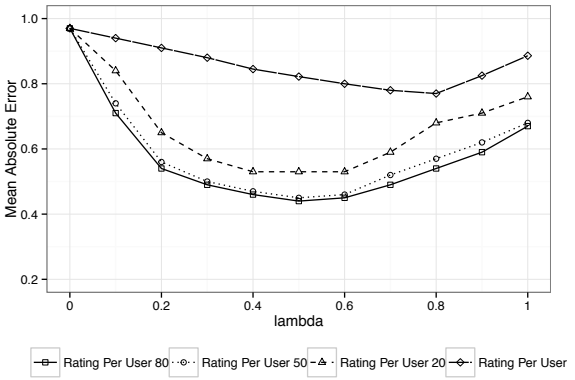
$$MAE = \frac{\sum_{x,y,z} |r_{x,y,z} - \hat{r}_{x,y,z}|}{L} \tag{4.19}$$

Recap the roles of two parameters, λ and δ present the dependence of recommendation results on the dataset US and IS. In order to impact of the parameters, we first test λ, setting δ to zero. Afterwards, we nail down the λ to test δ. Fig. 4.9(a) presents MAE of recommendation results according to varying λ from zero to one. The graph depicts the results based on 5, 20, 50 and 80 participants. More specifically, the best performances of the recommendation algorithm are obtained with λ between 0.4 and 0.6. For testing the δ, we set λ to 0.5 as the optimal value. Fig. 4.9(b) plots the impact of δ accordingly. Likewise, we also investigate the performance based on 5, 20, 50 and 80 participants. In Fig. 4.9(b), MAE increases, when *delta* is lower than 0.6 and higher than 0.8, which indicates $\delta = 0.7$ can guarantee a good recommendation results. Considering Fig. 4.9(a) and 4.9(b) jointly, the recommendation depends on λ more than δ. In other words, more relying on dataset IS improves the performance of recommendation results. In order to figure out how the number of participants influence the recommendation results, additional experiments has been shown in Fig. 4.9(c). It presents the optimal value of λ and δ varying when the size of participants is varied from 5 to 200. More specifically, the optimal value of λ and δ are fixed at 0.5 and 0.7 respectively, when the size of participants increases.

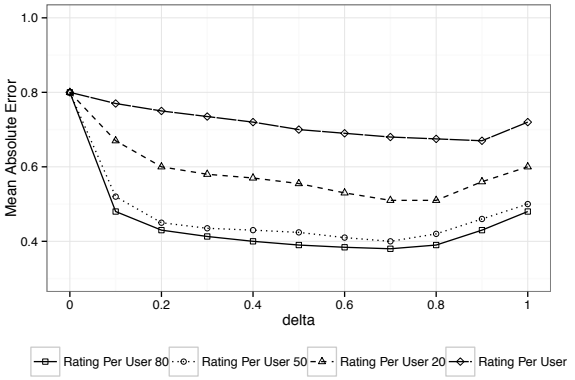
4.6 Discussion

In this section, we discuss some possible limitations of work, which may be argued.

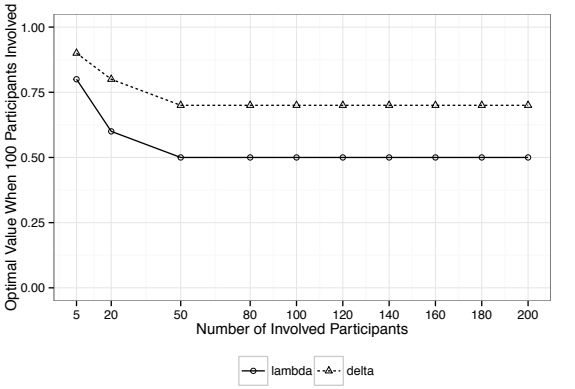
Firstly, we discuss the conception of privacy, since it is the foundation of our work. Nowadays, there is still no universally agreed-on definition of privacy in either research community or industry. We prefer to follow some prevalent interpretations. For example,



(a)



(b)



(c)

Fig. 4.9: Parameters estimation of the recommendation algorithm. (a) the impact of lambda (b) the impact of delta (c) the impact of size of participants

"privacy is private life, habits, act, relations and the right to be alone [WB90]" and "privacy is the claim of individuals, groups, or institutions to determine for themselves when, how, and to what extent information about them is communicated to others [Wes68]". Like we pointed in Section 4.2, these two acknowledged definitions both emphasized that privacy to people should be an ability to express themselves selectively. Therefore, we believe the privacy in the mobile is also based on users' expectations and decisions, especially in the current information age. It is very difficult to protect every single piece of data of each users, so the users have to express their expectations of privacy and we can help them to mitigate the corresponding privacy risk accordingly.

Secondly, we initialize the recommendation mechanism according to the collected users' privacy permission settings rather than the experts' opinions. This is something about our design philosophy, which is mentioned in Section 4.2. We consider there is no right answers for the people who want to set their privacy permission setting. In our case, we get some privacy permission settings from the participants during the experiments. Furthermore, the people are also allowed to make their choice on the privacy permission settings when they start to use our services. These two sources will be regarded as the initial dataset for generating the recommendations for each users.

Thirdly, we discuss the parameters in the recommendation approach in Section 4.5.3. We determined the parameters according to the MAE of recommendations. Also, according to our illustration, the number of participants also influences the performance of recommendation algorithms. The research issue about participant selection for generating recommendation algorithm is proposed, which is out of scope of this article and will be the future work.

Fourthly, there are more than 400 participants involved in our work to help us conduct the experiments and improve our research. We admitted that the more people participate,

the better the results will be. However, we cannot recruit as many participants as possible due to the time and resource limitation. Even though, we try our best to get more users involved. All the information about the participants are shown in Table 4.2. We avoid the statistical bias of the population, which can make our results more convincing.

Finally, there are two kinds of experiments to evaluate PriWe as shown in Section 4.5. One is based on Amazon Mechanical Turks, the other one is based on the real deployment. Both of them are based on the real users in the world. In the Amazon Mechanical Turks, we can get more participants in easily, which is significant to our work. In the real deployment, people will use our app and provide more feedbacks to us since we can have the face-to-face survey, which also can help us to improve our work. That is the reason why we conduct two sorts of evaluation.

4.7 Summary

In this chapter, we proposed PriWe, a system aims to understanding users' expectations of privacy and making recommendations about their privacy settings of installed mobile apps accordingly. We published a task on the Amazon Mechanical Turk and deployed PriWe in the real world for evaluation. According to the feedbacks of 382 participants from the Amazon Mechanical Turk, the recommendation made by PriWe can achieve around 78% accuracy for all the participants and achieve about 90% accuracy for the people in information privacy and security area. According to the feedbacks of 78 users from the real world, PriWe can make proper recommendations which can meet participants' privacy expectation and are mostly accepted by users, thereby help them to mitigate privacy disclosure in smartphone apps.

Chapter 5

Conclusions and Suggestions for Future Research

In this chapter, we conclude this thesis in Section 5.1 and outline some possible future works in Section 5.2.

5.1 Conclusions

Smartphone privacy is a significant issue in many fields including mobile computing, ubiquitous computing and internet of things. Much attention goes into this issue; many individuals and research communities devote themselves to providing the solutions. According to our survey in Chapter 2, most existing work can be viewed from two different perspectives, human-centric privacy and technology-centric privacy. Our works are mainly based on human-centric privacy since we believe the definition of privacy should depend on people's preference or attitude. In this thesis, we investigate smartphone privacy in mobile participatory sensing and mobile application. We mainly focus on privacy measurement and privacy mitigation in mobile computing. In each aspect, we identified the problems which lack sufficient studies and proposed corresponding solutions. We conclude these works as follows:

For privacy measurement in mobile participatory sensing, we proposed PriMe, a privacy

measurement method based on users' preferences towards data sharing in participatory sensing systems. Based on the proposed properties of privacy in participatory sensing, we measure the privacy according to individual attitude, which is represented by two intuitive properties: the inherent sensitivity of each data item; and the individual sensitivity to each data item. Experimental results illuminate that PriMe provides accurate results to the participants.

For mitigating privacy risk in mobile application, we proposed PriWe, a system aims to understanding users' expectations of privacy and making recommendations about their privacy settings of installed mobile apps accordingly. We published a task on the Amazon Mechanical Turk and deployed PriWe in the real world for evaluation. According to the feedbacks of 382 participants from the Amazon Mechanical Turk, the recommendation made by PriWe can achieve around 78% accuracy for all the participants and achieve about 90% accuracy for the people in information privacy and security area. According to the feedbacks of 78 users from the real world, PriWe can make proper recommendations which can meet participants' privacy expectation and are mostly accepted by users, thereby help them to mitigate privacy disclosure in smartphone apps.

In summary, smartphone privacy in mobile computing is a crucial issue. We have identified several important problems in different aspects of smartphone privacy, and proposed corresponding solutions. The evaluation results show that our approaches can mitigate risk and preserve privacy for users in the smartphone.

5.2 Suggestions for Future Research

We close this thesis by providing some suggestions for future research. Specifically, we believe that the following aspects are worth further investigations.

Challenge 1: Human Privacy and Smartphone Interaction. According to our discussion about human-centric smartphone privacy, a key challenge for the future is to build a system or framework based on interaction with human to protect privacy. As we emphasized in the Section 1.2, people' concerns are heterogeneous. Therefore, understanding people's privacy preference, concern and attitude should be an key challenge.

In the future, the smartphone privacy protection should be supported by concepts and methodologies issued from various disciplines, such as psychology, computer science and behavioristics.

Challenge 2: Active Defense in Smartphone Privacy. In comparison with human-centric privacy, active defense in smartphone privacy would focus on protecting privacy even without people's awareness and intervention. The active defense may collect users' behaviors data for learning and protect their information accordingly. Malicious apps currently have a plethora of ways to attack smartphones, even through an official application [DLZZ14].

To address such a problem, active defense technology is a future direction. It would involve privacy risk detection, privacy analysis, and protection. The adaptive version of active defense may also consider people's concerns and preferences.

Challenge 3: Smartphone Privacy Measurement and Analysis. Different criteria and metrics are currently being used to evaluate the performance of the proposed solutions in terms of privacy protection for different context [BGS11, PTSL13, MA12]. To achieve more precise and usable privacy-preserving in the smartphone, measurement and analysis for individual privacy in different environment should be proposed and applied. While it might be arduous or even impossible to propose an one-size-for-all measurement and analysis mechanism, the need to define generalized metrics is widely acknowledged.

Challenge 4: Smartphone Privacy Policy Modeling. It is not uncommon to realized that from scientific and technological viewpoints, there is no clear and absolute definition of privacy even there are some meaningful and acknowledged statements about privacy. A key challenge for the future is to propose a model as a unifying approach to formally state the smartphone privacy. The unified model can be not only a method to protect users' smartphone privacy but also a common metric to verified different algorithms, tools and systems. It even can be a reference to lawmakers when they legislate to protect citizen's information.

Bibliography

[ABK12] Sasikanth Avancha, Amit Baxi, and David Kotz. Privacy in mobile technology for personal healthcare. *ACM Computing Surveys (CSUR)*, 45(1):3, 2012.

[AFT+12] Domenico Amalfitano, Anna Rita Fasolino, Porfirio Tramontana, Salvatore De Carmine, and Atif M Memon. Using gui ripping for automated testing of android applications. In *Proceedings of the 27th IEEE/ACM International Conference on Automated Software Engineering*, pages 258–261. ACM, 2012.

[AH13] Yuvraj Agarwal and Malcolm Hall. Protectmyprivacy: detecting and mitigating privacy leaks on ios devices using crowdsourcing. In *Proceeding of the 11th annual international conference on Mobile systems, applications, and services*, pages 97–110. ACM, 2013.

[Alt77] Irwin Altman. Privacy regulation: culturally universal or culturally specific? *Journal of Social Issues*, 33(3):66–84, 1977.

[and] Android M Permissions: Best Practices for Developers. `https://developer.android.com/preview/features/runtime-permissions.html`.

[and14] Number of available android applications. `http://www.appbrain.com/stats/number-of-android-apps`, 2014.

[APN+14] Berker Agir, Thanasis G Papaioannou, Rammohan Narendula, Karl Aberer, and Jean-Pierre Hubaux. User-side adaptive protection of location privacy in participatory sensing. *GeoInformatica*, 18(1):165–191, 2014.

[app14] Your privacy is important to apple. So weve developed a Privacy Policy that covers how we collect, use, disclose, transfer, and store your information. `https://www.apple.com/legal/privacy/en-ww/`, 2014.

[BCMVO12] David Barrera, Jeremy Clark, Daniel McCarney, and Paul C Van Oorschot. Understanding and improving app installation security mechanisms through empirical analysis of android. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, pages 81–92. ACM, 2012.

[BD03] Louise Barkhuus and Anind K Dey. Location-based services for mobile telephony: a study of users' privacy concerns. In *INTERACT*, volume 3, pages 702–712. Citeseer, 2003.

[BDD+11] Sven Bugiel, Lucas Davi, Alexandra Dmitrienko, Stephan Heuser, Ahmad-Reza Sadeghi, and Bhargava Shastry. Practical and lightweight domain isolation on android. In *Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices*, pages 51–62. ACM, 2011.

[BF13] Trevor G Bond and Christine M Fox. *Applying the Rasch model: Fundamental measurement in the human sciences*. Psychology Press, 2013.

[BGR13] Zinaida Benenson, Freya Gassmann, and Lena Reinfelder. Android and ios users' differences concerning security and privacy. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 817–822. ACM, 2013.

[BGS11] Alex Braunstein, Laura Granka, and Jessica Staddon. Indirect content privacy surveys: measuring privacy without asking about it. In *Proceedings of the Seventh Symposium on Usable Privacy and Security*, page 15. ACM, 2011.

[BHS13] Sven Bugiel, Stephan Heuser, and Ahmad-Reza Sadeghi. Flexible and fine-grained mandatory access control on android for diverse security and privacy policies. In *Usenix security*, pages 131–146, 2013.

[BJL+13] Rebecca Balebako, Jaeyeon Jung, Wei Lu, Lorrie Faith Cranor, and Carolyn Nguyen. Little brothers watching you: Raising awareness of data leaks on smartphones. In *Proceedings of the Ninth Symposium on Usable Privacy and Security*, page 12. ACM, 2013.

[BRSS11] Alastair R Beresford, Andrew Rice, Nicholas Skehin, and Ripduman Sohan. Mockdroid: trading privacy for application functionality on smartphones. In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, pages 49–54. ACM, 2011.

[BS93] Victoria Bellotti and Abigail Sellen. Design for privacy in ubiquitous computing environments. In *Proceedings of the Third European Conference on Computer-Supported Cooperative Work 13–17 September 1993, Milan, Italy ECSCW93*, pages 77–92. Springer, 1993.

[BZNT11] Iker Burguera, Urko Zurutuza, and Simin Nadjm-Tehrani. Crowdroid: behavior-based malware detection system for android. In *Proceedings of the 1st ACM workshop on Security and privacy in smartphones and mobile devices*, pages 15–26. ACM, 2011.

[CCCS11] Haksoo Choi, Supriyo Chakraborty, Zainul M Charbiwala, and Mani B Srivastava. Sensorsafe: a framework for privacy-preserving management of personal sensory information. In *Secure Data Management*, pages 85–100. Springer, 2011.

[CFGW11] Erika Chin, Adrienne Porter Felt, Kate Greenwood, and David Wagner. Analyzing inter-application communication in android. In *Proceedings of the 9th international conference on Mobile systems, applications, and services*, pages 239–252. ACM, 2011.

[CGR+11] Delphine Christin, Julien Guillemet, Andreas Reinhardt, Matthias Hollick, and Salil S Kanhere. Privacy-preserving collaborative path hiding for participatory sensing applications. In *Mobile Adhoc and Sensor Systems (MASS), 2011 IEEE 8th International Conference on*, pages 341–350. IEEE, 2011.

[CJLF13] Eun Kyoung Choe, Jaeyeon Jung, Bongshin Lee, and Kristie Fisher. Nudging people away from privacy-invasive mobile apps through visual framing. In *Human-Computer Interaction–INTERACT 2013*, pages 74–91. Springer, 2013.

[CKK+08] Cory Cornelius, Apu Kapadia, David Kotz, Dan Peebles, Minho Shin, and Nikos Triandopoulos. Anonysense: privacy-aware people-centric sensing. In *Proceedings of the 6th international conference on Mobile systems, applications, and services*, pages 211–224. ACM, 2008.

[CLCC15] Paul Y Cao, Gang Li, Guoxing Chen, and Biao Chen. Mobile data collection frameworks: A survey. In *Proceedings of the 2015 Workshop on Mobile Big Data*, pages 25–30. ACM, 2015.

[COWC+13] Ningning Cheng, Xinlei Oscar Wang, Wei Cheng, Prasant Mohapatra, and Aruna Seneviratne. Characterizing privacy leakage of public wifi networks for users on travel. In *INFOCOM, 2013 Proceedings IEEE*, pages 2769–2777. IEEE, 2013.

[CRH+13] Delphine Christin, Christian Roßkopf, Matthias Hollick, Leonardo A Martucci, and Salil S Kanhere. Incognisense: An anonymity-preserving reputation framework for participatory sensing applications. *Pervasive and mobile Computing*, 9(3):353–371, 2013.

[CRKH11] Delphine Christin, Andreas Reinhardt, Salil S Kanhere, and Matthias Hollick. A survey on privacy in mobile participatory sensing applications. *Journal of Systems and Software*, 84(11):1928–1946, 2011.

[CS13] Andrew Clarke and Robert Steele. Health participatory sensing networks. *Mobile Information Systems*, 2013.

[DCS11] Emiliano De Cristofaro and Claudio Soriente. Short paper: Pepsi—privacy-enhanced participatory sensing infrastructure. In *Proceedings of the fourth ACM conference on Wireless network security*, pages 23–28. ACM, 2011.

[DLZZ14] Wenrui Diao, Xiangyu Liu, Zhe Zhou, and Kehuan Zhang. Your voice assistant is mine: How to abuse speakers to steal information and control your phone. In *Proceedings of the 4th ACM Workshop on Security and Privacy in Smartphones & Mobile Devices*, pages 63–74. ACM, 2014.

[DMC14] Quang Do, Ben Martini, and Kim-Kwang Raymond Choo. Enhancing user privacy on android mobile devices via permissions removal. In *System Sciences*

(HICSS), 2014 47th Hawaii International Conference on, pages 5070–5079. IEEE, 2014.

[EC12] Varick L Erickson and Alberto E Cerpa. Thermovote: participatory sensing for efficient building hvac conditioning. In *Proceedings of the Fourth ACM Workshop on Embedded Sensing Systems for Energy-Efficiency in Buildings*, pages 9–16. ACM, 2012.

[EGC+14] William Enck, Peter Gilbert, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. Taintdroid: an information flow tracking system for real-time privacy monitoring on smartphones. *Communications of the ACM*, 57(3):99–106, 2014.

[EKKV11] Manuel Egele, Christopher Kruegel, Engin Kirda, and Giovanni Vigna. Pios: Detecting privacy leaks in ios applications. In *NDSS*, 2011.

[FCH+11] Adrienne Porter Felt, Erika Chin, Steve Hanna, Dawn Song, and David Wagner. Android permissions demystified. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 627–638. ACM, 2011.

[FEW12] Adrienne Porter Felt, Serge Egelman, and David Wagner. I've got 99 problems, but vibration ain't one: a survey of smartphone users' concerns. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, pages 33–44. ACM, 2012.

[FHE+12] Adrienne Porter Felt, Elizabeth Ha, Serge Egelman, Ariel Haney, Erika Chin, and David Wagner. Android permissions: User attention, comprehension, and behavior. In *Proceedings of the Eighth Symposium on Usable Privacy and Security*, page 3. ACM, 2012.

[FLL+13] Bin Fu, Jialiu Lin, Lei Li, Christos Faloutsos, Jason Hong, and Norman Sadeh. Why people hate your app: Making sense of user feedback in a mobile app store. In *Proceedings of the 19th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 1276–1284. ACM, 2013.

[FSH12] Julien Freudiger, Reza Shokri, and Jean-Pierre Hubaux. Evaluating the privacy risk of location-based services. In *Financial Cryptography and Data Security*, pages 31–46. Springer, 2012.

[FZ94] George H. Forman and John Zahorjan. The challenges of mobile computing. *Computer*, 27(4):38–47, 1994.

[GB09] Victor Manuel García-Barrios. User-centric privacy framework: Integrating legal, technological and human aspects into user-adapting systems. In *Computational Science and Engineering, 2009. CSE'09. International Conference on*, volume 3, pages 176–181. IEEE, 2009.

[GCEC12] Clint Gibler, Jonathan Crussell, Jeremy Erickson, and Hao Chen. *AndroidLeaks: automatically detecting potential privacy leaks in android applications on a large scale*. Springer, 2012.

[GCJW10] Peter Gilbert, Landon P Cox, Jaeyeon Jung, and David Wetherall. Toward trustworthy mobile sensing. In *Proceedings of the Eleventh Workshop on Mobile Computing Systems & Applications*, pages 31–36. ACM, 2010.

[GEH+12] Michael M Groat, Benjamin Edwards, James Horey, Wenbo He, and Stephanie Forrest. Enhancing privacy in participatory sensing applications with multidimensional data. In *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, pages 144–152. IEEE, 2012.

[GJFJ12] Dibyajyoti Ghosh, Anupam Joshi, Tim Finin, and Pramod Jagtap. Privacy control in smart phones using semantically rich reasoning and context modeling. In *Security and Privacy Workshops (SPW), 2012 IEEE Symposium on*, pages 82–85. IEEE, 2012.

[GL05] Bugra Gedik and Ling Liu. Location privacy in mobile systems: A personalized anonymization model. In *Distributed Computing Systems, 2005. ICDCS 2005. Proceedings. 25th IEEE International Conference on*, pages 620–629. IEEE, 2005.

[GL08] Bugra Gedik and Ling Liu. Protecting location privacy with personalized k-anonymity: Architecture and algorithms. *Mobile Computing, IEEE Transactions on*, 7(1):1–18, 2008.

[GLP10] Nan-Wei Gong, Mathew Laibowitz, and Joseph A Paradiso. Dynamic privacy management in pervasive sensor networks. In *Ambient Intelligence*, pages 96–106. Springer, 2010.

[GPA+10] Raghu K Ganti, Nam Pham, Hossein Ahmadi, Saurabh Nangia, and Tarek F Abdelzaher. Greengps: a participatory sensing fuel-efficient maps application. In *Proceedings of the 8th international conference on Mobile systems, applications, and services*, pages 151–164. ACM, 2010.

[GPTA08] Raghu K Ganti, Nam Pham, Yu-En Tsai, and Tarek F Abdelzaher. Poolview: stream privacy for grassroots participatory sensing. In *Proceedings of the 6th ACM conference on Embedded network sensor systems*, pages 281–294. ACM, 2008.

[GZWJ12] Michael C Grace, Yajin Zhou, Zhi Wang, and Xuxian Jiang. Systematic detection of capability leaks in stock android smartphones. In *NDSS*, 2012.

[HHJ+11] Peter Hornyack, Seungyeop Han, Jaeyeon Jung, Stuart Schechter, and David Wetherall. These aren't the droids you're looking for: retrofitting android to protect data from imperious applications. In *Proceedings of the 18th ACM conference on Computer and communications security*, pages 639–652. ACM, 2011.

[HKBR99] Jonathan L Herlocker, Joseph A Konstan, Al Borchers, and John Riedl. An algorithmic framework for performing collaborative filtering. In *Proceedings of the 22nd annual international ACM SIGIR conference on Research and development in information retrieval*, pages 230–237. ACM, 1999.

[HMN+13] Shashank Holavanalli, Don Manuel, Vishwas Nanjundaswamy, Brian Rosenberg, Feng Shen, Steven Y Ko, and Lukasz Ziarek. Flow permissions for android. In *Automated Software Engineering (ASE), 2013 IEEE/ACM 28th International Conference on*, pages 652–657. IEEE, 2013.

[IAKR15] Qatrunnada Ismail, Tousif Ahmed, Apu Kapadia, and Michael K Reiter. Crowdsourced exploration of security configurations. In *Proceedings of the 33rd ACM Conference on Human Factors in Computing Systems*, 2015.

[JPBN09] Lukasz Jedrzejczyk, Blaine A Price, Arosha K Bandara, and Bashar Nuseibeh. I know what you did last summer: risks of location data leakage in mobile and social computing. *Department of Computing Faculty of Mathematics, Computing and Technology The Open University*, pages 1744–1986, 2009.

[JVLL12] Luis Gabriel Jaimes, Idalides Vergara-Laurens, and Miguel A Labrador. A location-based incentive mechanism for participatory sensing systems with budget constraints. In *Pervasive Computing and Communications (PerCom), 2012 IEEE International Conference on*, pages 103–108. IEEE, 2012.

[JZ13] Xuxian Jiang and Yajin Zhou. A survey of android malware. In *Android Malware*, pages 3–20. Springer, 2013.

[JZFF10] Dietmar Jannach, Markus Zanker, Alexander Felfernig, and Gerhard Friedrich. *Recommender systems: an introduction*. Cambridge University Press, 2010.

[Kan11] Salil S Kanhere. Participatory sensing: Crowdsourcing data from mobile smartphones in urban spaces. In *Mobile Data Management (MDM), 2011 12th IEEE International Conference on*, volume 2, pages 3–6. IEEE, 2011.

[KBSW13] Bastian Konings, Christoph Bachmaier, Florian Schaub, and Michael Weber. Device names in the wild: Investigating privacy risks of zero configuration networking. In *Mobile Data Management (MDM), 2013 IEEE 14th International Conference on*, volume 2, pages 51–56. IEEE, 2013.

[KCC+12] Patrick Gage Kelley, Sunny Consolvo, Lorrie Faith Cranor, Jaeyeon Jung, Norman Sadeh, and David Wetherall. A conundrum of permissions: installing applications on an android smartphone. In *Financial Cryptography and Data Security*, pages 68–79. Springer, 2012.

[KCS13] Patrick Gage Kelley, Lorrie Faith Cranor, and Norman Sadeh. Privacy as part of the app decision-making process. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 3393–3402. ACM, 2013.

[KKHK12] Denis Foo Kune, John Koelndorfer, Nicholas Hopper, and Yongdae Kim. Location leaks on the gsm air interface. *ISOC NDSS (Feb 2012)*, 2012.

[KS11a] Leyla Kazemi and Cyrus Shahabi. A privacy-aware framework for participatory sensing. *ACM SIGKDD Explorations Newsletter*, 13(1):43–51, 2011.

[KS11b] Leyla Kazemi and Cyrus Shahabi. Towards preserving privacy in participatory sensing. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2011 IEEE International Conference on*, pages 328–331. IEEE, 2011.

[KSK14] Ryoma Kawajiri, Masamichi Shimosaka, and Hisashi Kahima. Steered crowdsensing: incentive design towards quality-oriented place-centric crowdsensing. In *Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 691–701. ACM, 2014.

[KXAA13] Wazir Zada Khan, Yang Xiang, Mohammed Y Aalsalem, and Quratulain Arshad. Mobile phone sensing systems: A survey. *IEEE Communications Surveys & Tutorials*, 15(1):402–427, 2013.

[LAH+12] Jialiu Lin, Shahriyar Amini, Jason I Hong, Norman Sadeh, Janne Lindqvist, and Joy Zhang. Expectation and purpose: understanding users' mental models of mobile app privacy through crowdsourcing. In *Proceedings of the 2012 ACM Conference on Ubiquitous Computing*, pages 501–510. ACM, 2012.

[LBS+13] Jialiu Lin, Michael Benisch, Norman Sadeh, Jianwei Niu, Jason Hong, Banghui Lu, and Shaohui Guo. A comparative study of location-sharing privacy preferences in the united states and china. *Personal and ubiquitous computing*, 17(4):697–711, 2013.

[LCYZ15] Rui Liu, Jiannong Cao, Lei Yang, and Kehuan Zhang. Priwe: Recommendation for privacy settings of mobile apps based on crowdsourced users' expectations. In *2015 IEEE International Conference on Mobile Services (MS)*, pages 150–157, 2015.

[LD14] Bhushan Lokhande and Sunita Dhavale. Overview of information flow tracking techniques based on taint analysis for android. In *Computing for Sustainable Global Development (INDIACom), 2014 International Conference on*, pages 749–753. IEEE, 2014.

[LHX13] Hong Ping Li, Haibo Hu, and Jianliang Xu. Nearby friend alert: location anonymity in mobile geosocial networks. *Pervasive Computing, IEEE*, 12(4):62–70, 2013.

[LLV07] Ninghui Li, Tiancheng Li, and Suresh Venkatasubramanian. t-closeness: Privacy beyond k-anonymity and l-diversity. In *ICDE*, volume 7, pages 106–115, 2007.

[LML+10] Nicholas D Lane, Emiliano Miluzzo, Hong Lu, Daniel Peebles, Tanzeem Choudhury, and Andrew T Campbell. A survey of mobile phone sensing. *Communications Magazine, IEEE*, 48(9):140–150, 2010.

[LPA+11] Hieu Khac Le, Jeff Pasternack, Hossein Ahmadi, M Gupta, Y Sun, T Abdelzaher, Jiawei Han, Dan Roth, B Szymanski, and Sibel Adali. Apollo: Towards factfinding in participatory sensing. In *Information Processing in Sensor Networks (IPSN), 2011 10th International Conference on*, pages 129–130. IEEE, 2011.

[LYL+10] Hong Lu, Jun Yang, Zhigang Liu, Nicholas D Lane, Tanzeem Choudhury, and Andrew T Campbell. The jigsaw continuous sensing engine for mobile phone applications. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems*, pages 71–84. ACM, 2010.

[MA12] Norshidah Mohamed and Ili Hawa Ahmad. Information privacy concerns, antecedents and privacy measure use in social networking sites: Evidence from malaysia. *Computers in Human Behavior*, 28(6):2366–2375, 2012.

[MBK+12] Ildar Muslukhov, Yazan Boshmaf, Cynthia Kuo, Jonathan Lester, and Konstantin Beznosov. Understanding users' requirements for data protection in smartphones. In *Data Engineering Workshops (ICDEW), 2012 IEEE 28th International Conference on*, pages 228–235. IEEE, 2012.

[Min04] Robert P Minch. Privacy issues in location-aware mobile devices. In *System Sciences, 2004. Proceedings of the 37th Annual Hawaii International Conference on*, pages 10–pp. IEEE, 2004.

[MKGV07] Ashwin Machanavajjhala, Daniel Kifer, Johannes Gehrke, and Muthuramakrishnan Venkitasubramaniam. l-diversity: Privacy beyond k-anonymity. *ACM Transactions on Knowledge Discovery from Data (TKDD)*, 1(1):3, 2007.

[mos13] 12 Most Abused Android App Permissions. `http://about-threats.trendmicro.com/us/library/image-gallery/12-most-abused-android-app-permissions`, 2013.

[MPR08] Prashanth Mohan, Venkata N Padmanabhan, and Ramachandran Ramjee. Nericell: rich monitoring of road and traffic conditions using mobile smartphones. In *Proceedings of the 6th ACM conference on Embedded network*

sensor systems, pages 323–336. ACM, 2008.

[MS12] Christopher Mann and Artem Starostin. A framework for static detection of privacy leaks in android applications. In *Proceedings of the 27th Annual ACM Symposium on Applied Computing*, pages 1457–1462. ACM, 2012.

[MS14] Sandra R Murillo and J Alfredo Sánchez. Enhancing privacy awareness through interaction design. In *Proceedings of the XV International Conference on Human Computer Interaction*, page 44. ACM, 2014.

[MTG13] Alexios Mylonas, Marianthi Theoharidou, and Dimitris Gritzalis. Assessing privacy risks in android: a user-centric approach. In *Proceedings of the 1st international workshop on risk assessment and risk-driven testing (RISK-2013), Springer, Turkey (November 2013)*, 2013.

[Mul10] Collin Mulliner. Privacy leaks in mobile phone internet access. In *Intelligence in Next Generation Networks (ICIN), 2010 14th International Conference on*, pages 1–6. IEEE, 2010.

[Nis04] Helen Nissenbaum. Privacy as contextual integrity. *Wash. L. Rev.*, 79:119, 2004.

[OTK+12] Clemens Orthacker, Peter Teufl, Stefan Kraxberger, Günther Lackner, Michael Gissing, Alexander Marsalek, Johannes Leibetseder, and Oliver Prevenhueber. Android security permissions–can we trust them? In *Security and Privacy in Mobile Information and Communication Systems*, pages 40–51. Springer, 2012.

[PFB+14] Sebastian Poeplau, Yanick Fratantonio, Antonio Bianchi, Christopher Kruegel, and Giovanni Vigna. Execute this! analyzing unsafe and malicious

dynamic code loading in android applications. In *NDSS*, volume 14, pages 23–26, 2014.

[PHAB10] Georgios Portokalidis, Philip Homburg, Kostas Anagnostakis, and Herbert Bos. Paranoid android: versatile protection for smartphones. In *Proceedings of the 26th Annual Computer Security Applications Conference*, pages 347–356. ACM, 2010.

[PTSL13] Xinru Page, Karen Tang, Fred Stutzman, and Airi Lampinen. Measuring networked social privacy. In *Proceedings of the 2013 conference on Computer supported cooperative work companion*, pages 315–320. ACM, 2013.

[Ric06] John Rice. *Mathematical statistics and data analysis*. Cengage Learning, 2006.

[RQM13] Sanae Rosen, Zhiyun Qian, and Z Morely Mao. Appprofiler: a flexible method of exposing privacy-related behavior in android applications to end users. In *Proceedings of the third ACM conference on Data and application security and privacy*, pages 221–232. ACM, 2013.

[SCP+11] Minho Shin, Cory Cornelius, Dan Peebles, Apu Kapadia, David Kotz, and Nikos Triandopoulos. Anonysense: A system for anonymous opportunistic sensing. *Pervasive and Mobile Computing*, 7(1):16–30, 2011.

[SH14] Jialiu Lin Bin Liu Norman Sadeh and Jason I Hong. Modeling users mobile app privacy preferences: Restoring usability in a sea of permission settings. In *Symposium on Usable Privacy and Security (SOUPS)*, 2014.

[SHC+09] Norman Sadeh, Jason Hong, Lorrie Cranor, Ian Fette, Patrick Kelley, Madhu Prabaker, and Jinghai Rao. Understanding and capturing people's privacy

policies in a mobile social networking application. *Personal and Ubiquitous Computing*, 13(6):401–412, 2009.

[smm] Smartphone OS Market Share, Q3 2014. `http://www.idc.com/prodserv/smartphone-os-market-share.jsp`.

[SMSB14] Irina Shklovski, Scott D Mainwaring, Halla Hrund Skúladóttir, and Höskuldur Borgthorsson. Leakiness and creepiness in app space: perceptions of privacy and mobile app use. In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 2347–2356. ACM, 2014.

[sta14] The Statistics Portal. `https://www.statista.com/search/?statistics=1&studies=1&industryReports=1&dossiers=1&infos=1&subCategory=0&interval=0&category=0&subCategory=0&region=0&price=0&archive=0&q=how+many+apps+in+the+smartphone&sortMethod=idrelevance&accuracy=and&itemsPerPage=25`, 2014.

[Swe02] Latanya Sweeney. k-anonymity: A model for protecting privacy. *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, 10(05):557–570, 2002.

[SZL10] Jing Shi, Yanchao Zhang, and Yunzhong Liu. Prisense: privacy-preserving data aggregation in people-centric urban sensing systems. In *INFOCOM, 2010 Proceedings IEEE*, pages 1–9. IEEE, 2010.

[tan14] Android System Permissions. `http://developer.android.com/guide/topics/security/permissions.html`, 2014.

[Toc14] Eran Toch. Crowdsourcing privacy preferences in context-aware applications. *Personal and ubiquitous computing*, 18(1):129–141, 2014.

[top14] 92% of top 500 android apps carry security or privacy risk. http://www.infosecurity-magazine.com/news/92-of-top-500-android-apps-carry-security-or/, 2014.

[VKSW11] Michael Von Kaenel, Philipp Sommer, and Roger Wattenhofer. Ikarus: large-scale participatory sensing at high altitudes. In *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications*, pages 63–68. ACM, 2011.

[VVC11] Timothy Vidas, Daniel Votipka, and Nicolas Christin. All your droid are belong to us: A survey of current android attacks. In *WOOT*, pages 81–90, 2011.

[VZG12] Khuong Vu, Rong Zheng, and Jie Gao. Efficient algorithms for k-anonymous location privacy in participatory sensing. In *INFOCOM, 2012 Proceedings IEEE*, pages 2399–2407. IEEE, 2012.

[WB90] Samuel D Warren and Louis D Brandeis. The right to privacy. *Harvard law review*, 4(5):193–220, 1890.

[Wes68] Alan F Westin. Privacy and freedom. *Washington and Lee Law Review*, 25(1):166, 1968.

[Wic13] Stephen B Wicker. *Cellular Convergence and the Death of Privacy*. Oxford University Press, 2013.

[XLY+05] Gui-Rong Xue, Chenxi Lin, Qiang Yang, WenSi Xi, Hua-Jun Zeng, Yong Yu, and Zheng Chen. Scalable collaborative filtering using cluster-based smoothing. In *Proceedings of the 28th annual international ACM SIGIR conference*

on Research and development in information retrieval, pages 114–121. ACM, 2005.

[xpo] Xposed Module Repository. `http://repo.xposed.info/`.

[Yao95] YY Yao. Measuring retrieval effectiveness based on user preference of documents. *JASIS*, 46(2):133–145, 1995.

[YY12a] Lok-Kwong Yan and Heng Yin. Droidscope: Seamlessly reconstructing the os and dalvik semantic views for dynamic android malware analysis. In *USENIX Security Symposium*, pages 569–584, 2012.

[YY12b] Zhemin Yang and Min Yang. Leakminer: Detect information leakage on android with static taint analysis. In *Software Engineering (WCSE), 2012 Third World Congress on*, pages 101–104. IEEE, 2012.

[YYZ+13] Zhemin Yang, Min Yang, Yuan Zhang, Guofei Gu, Peng Ning, and X Sean Wang. Appintent: Analyzing sensitive data transmission in android for privacy leakage detection. In *Proceedings of the 2013 ACM SIGSAC conference on Computer & communications security*, pages 1043–1054. ACM, 2013.

[ZJ13] Yajin Zhou and Xuxian Jiang. Detecting passive content leaks and pollution in android applications. In *NDSS*, 2013.

[ZXGC14] Hengshu Zhu, Hui Xiong, Yong Ge, and Enhong Chen. Mobile app recommendations with security and privacy awareness. In *Proceedings of the 20th ACM SIGKDD international conference on Knowledge discovery and data mining*, pages 951–960. ACM, 2014.

[ZZD+12] Cong Zheng, Shixiong Zhu, Shuaifu Dai, Guofei Gu, Xiaorui Gong, Xinhui Han, and Wei Zou. Smartdroid: an automatic system for revealing ui-based

trigger conditions in android applications. In *Proceedings of the second ACM workshop on Security and privacy in smartphones and mobile devices*, pages 93–104. ACM, 2012.

[ZZJF11] Yajin Zhou, Xinwen Zhang, Xuxian Jiang, and Vincent W Freeh. Taming information-stealing smartphone applications (on android). In *Trust and Trustworthy Computing*, pages 93–107. Springer, 2011.