Privacy-Preserving Targeted Advertising for Mobile Devices

Yang Liu

Mansfield College

University of Oxford

A thesis submitted for the degree of

Doctor of Philosophy

Michaelmas Term 2017

Supervisor: Andrew Simpson

Acknowledgements

Six years ago, my friends and I started an IT company in Beijing, China. Our project collected and analysed mobile phone users' personal information in order to target them with specific advertisements. I have to admit that compared with protecting privacy, getting around its defenses was a faster way to make money at that stage. All is well if you can convince yourself that what you are doing is right. The question was, can I?

In 2014, I made a decision to leave the company, and to change my role from a data poacher to a privacy gamekeeper. Today, as an answer to myself, I finally complete this dissertation "Privacy-preserving Targeted Advertising for Mobile Devices", which is the product of three years of work here in Oxford. I could never complete this alone without the help of many, many people. I would like to thank each and every one of them for their support and company during my wonderful journey.

First and foremost, enormous gratitude is due to my supervisor, Andrew Simpson. Doctoral students often talk about how their supervisors influence the experience of "doing a PhD". In this respect, it has been an extremely pleasant experience to work with Andy. His style of guidance is very conscientious — his suggestions are always more than enough to inspire or direct me to the right source to solve problems, and yet are never so much that they change my research ideas or damage my confidence. Academia aside, he would tell me that rest is important, to accompany my family and do sports, as well as drive me out of the office when I tended to put a bit too much pressure on myself. When a paper was rejected with some "negative" comments he would encourage me with reminders that "such is life". In addition, he would say "the bill is on me" at every termly meet-up (in a random restaurant with all his students) — that is the most Chinese-style thing I experienced here. After more than three years, my respectful mentor has become a solid friend. I guess I can only use the word "fortunate" to express the feeling of having Andy's support all these years. Andy, thank you so much!

My thanks also go to Andrew Martin, Hamed Haddadi, and Kasper Rasmussen, for being my major advisors and assessors through my transfer, confirmation, and the final vivas. To be honest, I was quite nervous during the first viva. Fortunately, they set a great tone for it and I suddenly realised that the viva was not some test

designed to create difficulties, instead it was an excellent opportunity to discuss my work with experts and make it better. Their great suggestions have helped to improve the document you are reading.

Special mention goes to Cas Cremers, for being a second familiar professorial face, other than Andy. During the past three years, I have always seen him as a shadow member and a strong backing of our office. I also thank Jim Davies for his lively teaching of the CDS course and giving me a score of 83 at the end of it — I really got a lot of confidence from that.

Of course, I must thank my office mates: Katriel "the fencing master" Cohn-Gordon, Kevin "the Canadian secret agent" Milner, Chad "the bearded daddy" Heitzenrater, Martin "the 5th man of room 445" Dehnel-Wild. Their company has been invaluable over the years. Thanks also go to my colleagues who supported me in whatever shape or form: Alan Abe, Majed Alshammari, Robin Ankele, Aaron Ceross, Eduardo dos Santos, Jaco Jacobs, Yishu Miao, Norbert Nthala, Emma Osborn, Andrew Paverd, Zhenghua Xu.

I am indebted to the China Scholarship Council (CSC), the Department of Computer Science, Mansfield College and Oxford Kendo Club — to CSC for providing financial support; to my dear department for providing a high quality environment (and delicious coffee); to my lovely college for providing all the events and opportunities that helped me to experience British culture; and finally to Oxford Kendo Club for providing me the bruises, inspiration and vigorous spirit.

I must thank two special friends and give them my best wishes. Robert Ryan, the first friend I made in the UK, who knows more Chinese classical literature than I do. He really needs to drink less when he is not reading (he declared that he was wronged). Heidi Radburn, who provided so much help to my wife and I when we had just arrived in this unfamiliar country. Heidi, you are so close to success now. You can definitely beat the cancer. I am sure of that.

Lastly, but of course by no means least, my love and gratitude goes out to my family. A big thank you to my mom, Hua, who has taught me to love people and the world. A big thank you to my dad, Feng, who has taught me to respect others and to take responsibility. A big thank you to my parents-in-law, Yeyun and Xiaozhen, who trust me enough to let their baby daughter come and live in a distant country with me. Special mention also goes to the special members of my family: Puppy Niuniu (1999—2012) and Puppy Pidan (saved from the cold street in 2013), their innocent faces have comforted since I was a kid. A special thank you to Henan, the best of wives and the best of women, I could never be who I am today without her extraordinary support. We are expecting our first child in late May, 2018. I dedicate the thesis to the hope of our future — we will try to make a better world for you.

Abstract

With the continued proliferation of mobile devices, the collection of information associated with such devices and their users — such as location, installed applications and cookies associated with built-in browsers — has become increasingly straightforward. By analysing such information, organisations are often able to deliver more relevant and better focused advertisements. Although such *Targeted Mobile Advertising* (TMA) offers great benefits to advertisers, it gives rise to a number of concerns, with privacy-related concerns being prominent amongst them. It follows that there is a need for an advertisement-selection mechanism that can support the existing TMA business model in a manner that takes into account consumers' privacy concerns.

The research described in this dissertation explores the delicate balance between the goals of the advertisers and the consumers: advertisers pursue profits by applying TMA, which violates consumers' privacy; consumers hope to benefit from useful mobile advertisements without compromising their personal information. The conflicts of interests between consumers and advertisers in the context of targeted mobile advertising brings us to our research question: *Is it possible to develop a privacy-preserving TMA framework that enables mobile users to take advantage of useful advertising services without their privacy being compromised, and without impacting significantly advertising effectiveness?*

In order to answer this question, this dissertation presents four main contributions. First, we report upon the result of a qualitative study to discuss the balance that needs to be struck between privacy and utility in this emerging area. Second, a number of formal models are developed to reason about privacy, as well as to reason about the relationship between privacy and utility in the context of TMA. Third, a novel ad-selection architecture, *PPTMA* (Privacy-Preserving Targeted Mobile Advertising), is presented and evaluated. Finally, a privacy-preserving advertisement-selection mechanism, *AdSelector*, is introduced. The mechanism is novel in its combination of a user subscription mechanism, a two-stage ad-selection process, and the application of a trustworthy billing system.

Contents

List of Figures **xi**

List of Tables **xiii**

1 Introduction **1**
1.1 Motivation 1
1.2 Research problem and contributions 4
1.3 Papers arising from this work 6
1.4 Dissertation structure 6

2 Background **9**
2.1 The TMA ecosystem 9
2.2 Characteristics of, and privacy in, mobile advertising 11
2.2.1 Characteristics of mobile advertising 11
2.2.2 Privacy in mobile advertising 13
2.3 Related Work 17
2.3.1 Previous research on TMA 17
2.3.2 Previous research on privacy-preservation approaches 20
2.4 Summary 30

3 Perspectives on privacy **31**
3.1 The trade-off between privacy and utility 32
3.1.1 The unavoidable trade-off 32
3.1.2 The underlying free app ecosystem 33
3.1.3 Attitude towards the trade-off 34
3.1.4 The privacy paradox 36
3.1.5 The authenticity of collected information and the lack of privacy awareness 37
3.1.6 A real-life example 38
3.2 Human factors contributing to the trade-off: a user study 43
3.2.1 Overview of the user study 43
3.2.2 Study approaches and participants 45
3.2.3 Scenarios and data collection 46

3.2.4 Results and discussion 49
3.2.5 Lessons learned from the user study 59
3.3 Requirements and threats 62
3.3.1 Scope 62
3.3.2 High level requirements 63
3.3.3 Threat models 65
3.4 Summary 66

4 Formal models: From TMA to PPTMA 69
4.1 Formal models in the context of TMA 70
4.1.1 Formal models and privacy 70
4.1.2 Purposes of the formal models 71
4.2 From TMA to PPTMA 73
4.2.1 An abstract TMA system 73
4.2.2 PPTMA: A privacy-preserving solution 74
4.3 A model of TMA 75
4.3.1 Overview of the model 76
4.3.2 Privacy-affected operations 81
4.4 A model of PPTMA 85
4.4.1 Overview of the model 85
4.4.2 Improved operations 89
4.5 Application of the PPTMA model 92
4.5.1 The first stage of the ad-selection process: pre-download ads 93
4.5.2 The second stage of the ad-selection process: local ad selection 95
4.5.3 Click-audit obfuscating and click-fraud detecting 96
4.6 Analysis 98
4.7 Summary 100

5 PPTMA: A framework for privacy-preserving targeted mobile advertising 101
5.1 Design guidelines 102
5.2 System architecture 104
5.3 Prototype implementation 107
5.3.1 Ad scanning 107
5.3.2 Permissions and personal profile management 110
5.3.3 Sandbox 113
5.3.4 Simulation and evaluation 115
5.3.5 Privacy improvements introduced by PPTMA 118
5.4 Summary 120
5.4.1 Contribution over previous solutions 121
5.4.2 Possible objections 122

6 AdSelector: A Privacy-Preserving Advertisement Selection Mechanism for Mobile Devices 127
6.1 Overview of AdSelector . . . 128
6.1.1 Coordinate with PPTMA . . . 128
6.1.2 Ad interest categorisation in AdSelector . . . 130
6.2 System architecture . . . 131
6.3 Core mechanisms and workflow . . . 134
6.3.1 User subscription mechanism . . . 137
6.3.2 Two-stage ad selection process . . . 139
6.3.3 Privacy-conscious billing . . . 144
6.3.4 Advertising fraud defence . . . 146
6.4 Simulation and evaluation . . . 150
6.4.1 Simulation setup . . . 150
6.4.2 Simulation scenarios . . . 151
6.4.3 Privacy improvements introduced by Ad-Selector . . . 156
6.4.4 Evaluation of performance . . . 156
6.4.5 Limitations and challenges . . . 159
6.5 Summary . . . 160

7 A Roadmap 163
7.1 Economic models of incentives . . . 163
7.2 Potential application to other domains . . . 166
7.3 Connection with wider academic areas . . . 167
7.3.1 Internet privacy . . . 168
7.3.2 Economics of privacy . . . 168
7.4 Evolution of the TMA ecosystem . . . 169
7.5 Summary . . . 170

8 Conclusion 171
8.1 A review . . . 172
8.1.1 The user study: perspectives on privacy . . . 172
8.1.2 Formal models . . . 173
8.1.3 PPTMA . . . 173
8.1.4 AdSelector . . . 175
8.2 Measuring success against threats . . . 176
8.3 Future work . . . 177
8.3.1 Improve the framework . . . 177
8.3.2 Improve and extend the formal models . . . 178
8.3.3 Extend the framework to wider domains . . . 178
8.3.4 Research TMA from another perspective . . . 179
8.4 Outlook . . . 179

Appendices

A Questionnaire used in the qualitative study **183**

B Slides used in focus group sessions of the qualitative study **191**

References **197**

List of Figures

1.1 Structure of the dissertation 7

2.1 The TMA ecosystem 10

3.1 Facebook permissions list 39
3.2 Permission details 39
3.3 Permission lists of three weather apps 40
3.4 An app with targeted advertisements 41

4.1 A typical TMA workflow 73
4.2 The composition of the PPTMA system 85

5.1 PPTMA architecture 104
5.2 Ad scanning module 110
5.3 Permissions management module 112
5.4 Personal profile management module 114
5.5 Personal information process flow 119

6.1 Architecture overview of PPTMA 129
6.2 System architecture 132
6.3 Alternative system architecture 133
6.4 System core workflow 136
6.5 The two-stage ad selection process 141
6.6 Flow of steps in processing the view/click reports 145

List of Tables

3.1 Android distribution numbers for August 8, 2017 42
3.2 Participants description 46
3.3 Auxiliary data: Awareness and knowledge of privacy leakage 51
3.4 Auxiliary data: Impact of trade-off situations 53
3.5 Auxiliary data: Trust in the privacy-preserving technologies 58
3.6 factors influencing decision-making pertaining to the trade-off . . . 59
3.7 Threats 67

4.1 Analysis on released personal information and related effects: an example 98

5.1 Different ad styles supported by *Domob* 108
5.2 Software testing environment 115
5.3 Cost of the prototype 116
5.4 Compatibility testing 118

6.1 Time cost in Scenario 1 153
6.2 Selected ad and disclosed personal information in Scenario 1 153
6.3 Time cost in Scenario 2 154
6.4 Selected ads and disclosed personal information in Scenario 2 155
6.5 Consumption of storage and bandwidth of different ad formats . . . 157
6.6 Consumption of memory and energy (PPTMA + AdSelector) . . . 159

7.1 Stakeholders' inputs and outputs without PPTMA 164
7.2 Changes imported by PPTMA 165

1 Introduction

Contents

1.1 Motivation 1
1.2 Research problem and contributions 4
1.3 Papers arising from this work 6
1.4 Dissertation structure 6

1.1 Motivation

Online Targeted Advertising (OTA), also known as *Online Behavioural Advertising* (OBA), has brought significant benefits to advertisers since it was first proposed in the 1990s [1]. Although OTA leverages some relatively new technologies, the concept of 'behavioural advertising' is not new: marketers have collected customers' data to tailor advertising messages to specific audiences since at least the 1960s [2]. In the context of online commerce, DoubleClick began to track and analyse users' browsing patterns using third-party cookies for the purposes of presenting targeted advertisements in the late 1990s [3]. Using the information collected from users' online behaviour, including historical search queries, site views and other operations on particular types of content, advertisers are able to deliver the most relevant

advertisements (ads) to their users. These ads can match users' interests with a high level of accuracy; as such, OTA significantly improves advertising effectiveness.

To estimate how much help OTA can offer to advertisers, Yan *et al.* [4] reported on a study pertaining to the click-through log of advertisements collected from a commercial search engine. Their findings suggested that the *click-through rate* (CTR) of a particular advertisement can, on average, be improved by as much as 670% after segmenting users via behavioural targeting strategies. In addition, their study indicated that the CTR of ads could be further improved significantly — beyond 1,000% — with more advanced behavioural targeting strategies. Another survey [5] of 12 advertising networks in 2009 indicated that the average quarterly data on pricing for OTA is over twice that for standard run-of-network advertising. Goldfarb and Tucker [6] have suggested that advertising effectiveness decreased by about 65% in the EU relative to the rest of the world after the implementation of the ePrivacy Directive[1] (which limited the ability of advertisers to use and collect personal data for OTA) was passed. The benefits of OTA are, therefore, clear.

As one of the fastest growing segments of the OTA market, *Targeted Mobile Advertising* (TMA) has seen significant developments in recent years. In 2011, the value of mobile ad spending in the USA was approximately US $1.1 billion per year [7]; this figure reached US $30.45 billion by the end of 2015, accounting for 51.9% of total digital ad spending in the USA [8] — meaning that spending associated with mobile ads in the USA has now surpassed spending associated with desktop ads. In the ad spending report provided by eMarketer [9], it was predicted that the value of mobile ad spending in the UK would overtake that of TV ad spending and reach £4.58 billion per year, accounting for 50.8% of all digital ad spending and 27% of total media spending in the UK by the end of 2016 — meaning that desktop ad spending in the UK would be overtaken by mobile ad spending by the end of 2016.

In order to apply TMA, a new ecosystem has gradually been formed. Ad-networks bill advertisers for delivering ads, and share the payment with developers who provide space for displaying ads. In this mobile app ecosystem, many enterprises

[1] `http://eur-lex.europa.eu/LexUriServ/LexUriServ.do?uri=OJ:L:2002:201:0037:0047:en:PDF`

or developers provide their mobile apps for free and rely on targeted advertising for revenue. For example, according to a study from 2012 [10], about 80% of the free apps in the Google Play Market rely on mobile advertising as their main source of revenue. Our research [11] found that at least one ad library was embedded in 46 out of 60 off-the-shelf apps selected randomly from the top free apps of the Google Play Store in September 2015.

However, it is well known that mobile users typically cannot use the 'free apps' freely. In exchange for the benefits of the received services, users, who perhaps do not fully comprehend the consequences of their actions, voluntarily give up some of their privacy. To improve advertising effectiveness, ad-networks select tailor-made ads targeted at individuals by analysing consumers' personal information, which is primarily collected from consumers' mobile devices. Although TMA promises an improvement over traditional delivery approaches, users' privacy concerns, in turn, give rise to a degree of hostility towards advertisers, which might lead to the utilisation of ad-blocking tools. A number of studies (see, for example, the contributions of [12] and [13]) indicate a direct relationship between consumer attitudes and behaviour: consumers' trust in privacy-preservation of mobile advertising is positively related to their willingness to accept mobile advertising, and they generally have negative attitudes towards mobile advertising unless they have specifically consented to it.

The relationship between consumers and advertisers is a delicate one and needs to be handled appropriately: a solution to the privacy-preserving targeted mobile advertising 'problem' should provide benefits both to consumers — so that they can take advantage of ads without worrying about their privacy being compromised — and to the companies — resulting in reduced hostility and increased response rates. Previous work in this area has tended to take one 'side' or the other: those on the 'side' of consumers might propose solutions that deceive advertisers (e.g. [6] and [14]); those on the 'side' of companies might propose solutions that bypass privacy concerns (e.g. [15] and [16]).

Clearly, there is a need for a targeted mobile ad-selection mechanism that can support the existing TMA business model, and can do so in a manner that helps to protect users' personal information rather than simply bypassing their privacy concerns. Hence, we aim to develop a solution that strikes a balance between concerns and has the potential to be palatable to both 'sides'.

1.2 Research problem and contributions

TMA improves advertising effectiveness significantly for advertisers, and gives rise to a number of concerns on behalf of users, with privacy-related concerns being high on that list [17]. The research described in this dissertation explores the delicate balance between the goals of the advertisers and the consumers: advertisers pursue profits by applying TMA, which violates consumers' privacy; consumers hope to benefit from useful mobile advertisements without compromising their personal information. The conflicts of interests between consumers and advertisers in the context of targeted mobile advertising brings us to our research question:

> *Is it possible to develop a privacy-preserving TMA framework that enables mobile users to take advantage of useful advertising services without their privacy being compromised, and without impacting significantly advertising effectiveness?*

Driven by the research question, the dissertation presents four contributions:

1. The result of a user study that investigates the factors influencing decision-making pertaining to the trade-off between privacy and utility in mobile services. The results show users' general perspectives on privacy and provide potential privacy requirements for a privacy-preserving TMA system. The survey also explores directions that might be followed to foster adoption of privacy-preserving technologies to aid mobile users in balancing privacy and utility.

2. Formal models that describe the existing TMA system and a potential privacy-preserving TMA system. The formal models help in reasoning about privacy,

as well as in reasoning about the relationship between privacy and utility in the context of TMA. Privacy requirements drawn from the user study are also tested in the models to underpin our proposed privacy-preserving framework.

3. The design of a new ad-selection architecture, namely *PPTMA* (Privacy-Preserving Targeted Mobile Advertising). The solution tries to achieve a balance: it can be deployed by users, and is also intended to be palatable to businesses that operate in this space. This feature is enabled by supporting a configurable ad-selection mechanism module that can be customised by advertisers. An initial prototype of PPTMA is implemented on the Android system.

4. A privacy-preserving advertisement-selection mechanism, namely *AdSelector*. The mechanism is novel in its combination of a user subscription mechanism, a two-stage ad-selection process, and the application of a trustworthy billing system. In particular: (1) the user subscription mechanism helps users to identify their interests and subscribe to desirable categories of ads; (2) the two-stage ad-selection process ensures that ad servers can only obtain coarse-grained user profiles, with fine-grained user profiles stored and used only on the mobile devices; and (3) the trustworthy billing system helps to report ad-clicks without revealing users' identities and assists in detecting click-fraud attacks (a type of fraud that occurs in pay-per-click online advertising when the attackers click on an ad — rarely by hand, often by automated scripts — without actual interest in the target of the ad's link for the purpose of earning extra income or depleting competitors' advertising budgets [18]).

1.3 Papers arising from this work

The research described in the dissertation has contributed to the following publications (in the order of their appearance in this dissertation):

1. "On the trade-off between privacy and utility in mobile services: A qualitative study" (with A. C. Simpson). Submitted to the 14th Symposium on Usable Privacy and Security (SOUPS 2018).

2. "Privacy-preserving targeted mobile advertising: Formal models and analysis" (with A. C. Simpson). Proceedings of the 11th International Workshop on Data Privacy Management (DPM 2016). Lecture Notes in Computer Science Volume 9963, Pages 94—110, 2016.

3. "Privacy-preserving targeted mobile advertising: Requirements, design, and a prototype implementation" (with A. C. Simpson). Software: Practice and Experience. 46(12): 1657—1684, 2016.

4. "AdSelector: A privacy-preserving advertisement-selection mechanism for mobile devices" (with A. C. Simpson). The Computer Journal. 60(8): 1251—1270, 2017.

A visual representation of mappings between research contributions and papers is shown in Figure 1.1.

1.4 Dissertation structure

The reminder of the dissertation is organised as follows (see Figure 1.1).

Chapter 2 presents background information of the TMA systems. It describes the current state-of-the-art research in the field of targeted advertising. Chapters 3 – 6 represent the four contributions of this dissertation. A road map that discusses the considerations emerged from our work is summarised in Chapter 7. We then present conclusions in Chapter 8.

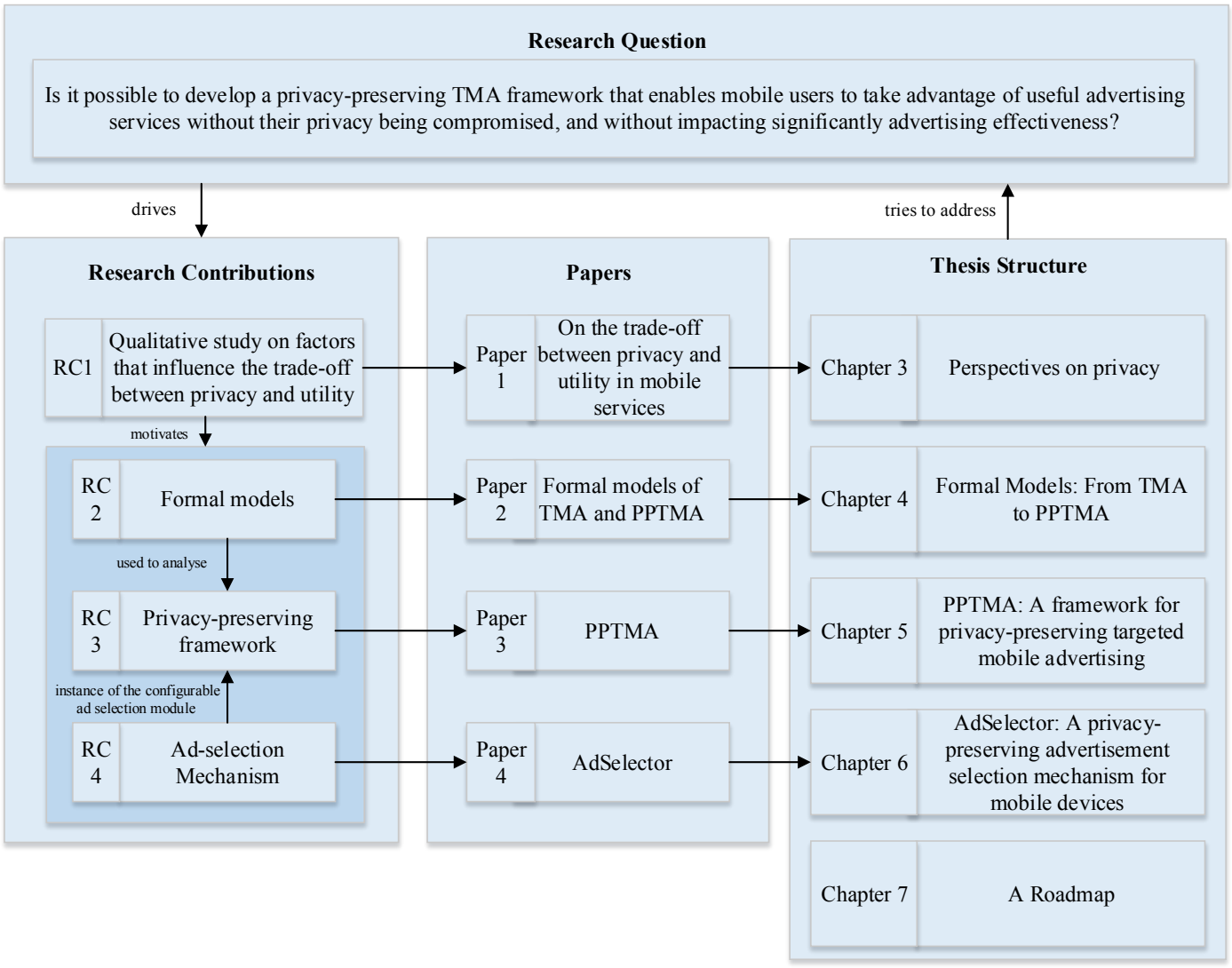

Figure 1.1: Structure of the dissertation

2 Background

Contents

2.1 The TMA ecosystem 9
2.2 Characteristics of, and privacy in, mobile advertising . 11
2.2.1 Characteristics of mobile advertising 11
2.2.2 Privacy in mobile advertising 13
2.3 Related Work 17
2.3.1 Previous research on TMA 17
2.3.2 Previous research on privacy-preservation approaches . 20
2.4 Summary 30

This chapter provides background information about the *Targeted Mobile Advertising (TMA)* ecosystem, and helps to motivate the contributions of this dissertation. Section 2.1 gives an overview of the current TMA ecosystem. Section 2.2 then discusses the characteristics of mobile advertising, as well as the different kinds of personal data stored in mobile devices. Section 2.3 reviews previous research in the context of TMA.

2.1 The TMA ecosystem

The core concept of a TMA system is the mechanism that can automatically select and display relevant ads for mobile users. In general, the automatic process is

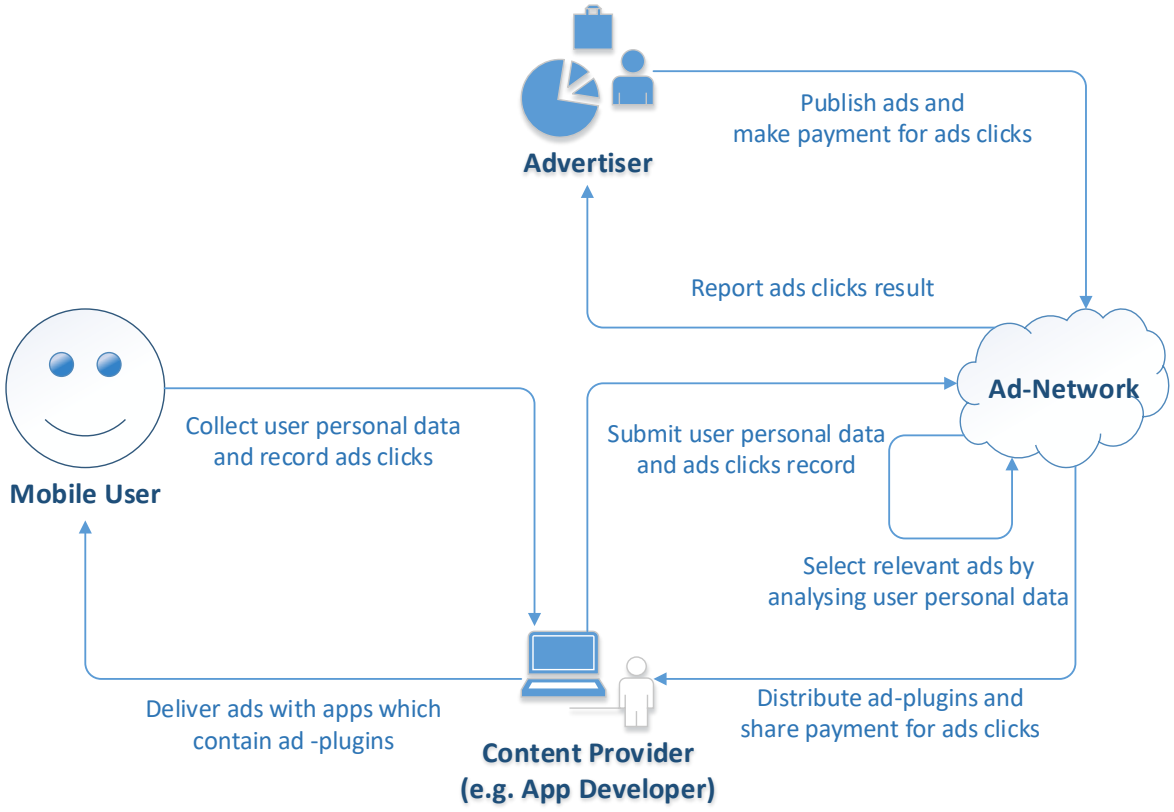

Figure 2.1: The TMA ecosystem

based on the user's data, which includes personal information, a record of their online behaviour, the current context they are in, and so on. For example, for a user who has installed a sports-related app on their mobile, and who frequently searches sports-related websites, a TMA system could present an ad of a sports store close to the user's current location.

There are four main kinds of participating stakeholders in the TMA ecosystem: advertisers; content providers; ad-networks; and users. We consider each in turn.

1. **Advertisers** are the people who own companies, shops, firms, and so on. They want to display their adverts on web pages or mobile apps to advertise their products and they are willing to pay for the advertising services.

2. **Content providers** own web sites or mobile apps and are willing to leave a space on their sites or apps to place ads for advertisers. By providing content resources they obtain corresponding payment.

3. **Ad-networks** collect ads from advertisers and serve them to content providers. They work as mediators between advertisers and content providers.

4. **Users** interact with their mobile devices, click ads they are interested in, and (typically) wish to retain their privacy.

Most ad-networks maintain their own TMA system, which consists of numerous apps and ads. Mobile users' personal data is collected by the ad-plugins that are embedded in the apps and sent to the ad-networks' servers. An ad-selection mechanism then selects ads from a large pool of ads submitted by advertisers. Advertisers need to pay related ad-networks if their ads are viewed or clicked by the users. The ad-networks will then share the payment with app developers (content providers) who display the ads within their apps. Figure 2.1 shows the interactions between these stakeholders.

2.2 Characteristics of, and privacy in, mobile advertising

Because of the unique features of, and the huge amount of personal information stored in, mobile devices, an efficient way of collecting personal data is provided. We started by discussing the characteristics of mobile advertising.

2.2.1 Characteristics of mobile advertising

Compared to web advertising or traditional advertising, advertisers are able to reach their target audience in a more efficient way with mobile advertising. We now discuss the special features of mobile devices that make it unique.

1. **Timeliness and portability**. Due to the nature of the devices, smartphone users can be targeted with mobile advertising services anytime and anywhere. Meanwhile, the ad-networks can also learn personal information by analysing data collected during specific periods of time and at specific locations to, for example, infer the user's place of work and home location.

2. **Fragmented time effect**. Individuals often take advantage of the time available when waiting in line or travelling. *Fragmented time* can be spent on some simple transactions, such as browsing news or checking emails. A large number of apps have been developed with mobile advertising targeting fragmented time in mind.

3. **Various sensors**. Sensors such as GPS and gyroscopes are built inside smartphones. Via utilisation of these sensors, interesting advertising solutions can be explored. For instance, Location Based Services (LBS) and Augmented Reality (AR) technology can be used together to deliver interactive ads for customers in special stores.

4. **Apps as advertising media**. Being limited by screen size, ads displayed in mobile browsers are not as conspicuous as those presented in PC browsers. In addition, due to one app being displayed per screen, ads inside apps are increasing in prevalence.

5. **Unique user**. Unlike, say, PCs, a mobile device is typically associated with a single user. However, the possibility of cross-device and coerced access [14] (for example, parents checking their children's phones) needs to be considered.

6. **Permanent identifier**. In the PC context, cookies are the most common way of associating an identifier with a user. However, cookies can be cleared from browser memory, and cookies of different browsers cannot interact with each other. By contrast, in the TMA context, advertisers can use device identifiers to track users — with these identifiers rarely changing.

7. **Viral marketing**. The viral marketing feature of mobile advertising can maximise the influence of ads with mobile users' initiating actions. For example, it may be that people will have less resistance to ads recommended by friends.

8. **More personal information**. Advertisers can obtain a preliminary profile of consumption capability on the basis of phone model. Users can be categorised into audience segments easily by analysing the categories of installed apps. Potential viral marketing targets can be located via address books and call logs. Users' various privacy contexts make it more effective for advertisers to perform TMA.

2.2.2 Privacy in mobile advertising

Although there are many different meanings in the term *privacy*, this thesis mainly focuses on the user-understandable aspects of privacy that are relevant to TMA systems. In the research of privacy decision making conducted by Acquisti and Grossklags [19], more than 90% of respondents very much agrees with the definition of privacy as:

> *Ownership and control of personal information.*

A similar definition of *privacy*, which is specifically relevant to TMA systems, is adopted in this thesis:

> *Ownership and control of personal information (and operations that can be used to infer that personal information) stored in users' own mobile devices.*

For instance, an app installed on a user's mobile device is a part of her privacy, as the category of such app might indicate the user's possible interest. Meanwhile, the operation of uninstalling an app is also a part of her privacy, as the operation could, in a certain extent, indicate that the user might be not interested in a related category of mobile service anymore.

Correspondingly, we define *privacy leakage* in this thesis as follows:

> *All or part of the personal information that is stored in users' own mobile devices is shared with any third-party other than the third-party to which the personal information was initially provided by users themselves.*

We now consider some of the most sensitive types of personal information stored in mobile devices. This information is the main resource that advertisers or third-parties might try to obtain. It is initially classified into two categories: dynamic information and static information. We consider each in turn.

Dynamic information represents user data that is generated and changed dynamically based on users' current behaviour or operations. Some examples are shown below.

1. **Location information.** Compared with OTA on PCs, a significant advantage of TMA is the effective utilisation of consumers' location information. A modern smartphone, as a portable item with various sensors and network resources, is an excellent tool to record the user's location information. Mobile apps can obtain users' precise locations using GPS or network location sources such as WiFi and mobile towers. The real-time location has the potential to enable advertisers to find consumers at particular times, for example delivering an advertisement for the nearest pizza restaurant just before lunchtime. Moreover, the record of past locations has the potential to help the advertisers discover where the user lives, where the user works, and the pattern of the user's daily routine.

2. **Clickstream.** By computing the clickstream (a list of URLs belonging to ad-networks visited by a user), advertisers can obtain users' behavioural profiles for performing TMA.

3. **Operations on apps.** Since all apps are assigned to explicit categories before they are uploaded to app stores, installations and removals of apps can expose the degree of a user's interest in particular areas. To quote the Director of digital marketing firm iCrossing, Rachel Pasqua, "Apple knows what you've downloaded, how much time you spend interacting with an app and even knows what you've downloaded over time and didn't like or deleted" [20].

Static information represents data that is bound to the hardware, or data that is abstracted from the history of users' operations. Some examples are shown below.

1. **Hardware information.** Ad-networks can obtain the hardware information of mobile devices via APIs provided by companies such as Google and Apple. This hardware information includes data pertaining to device type, phone model, screen size, and camera performance. By analysing this basic information, advertisers can create some initial profiles for their customers. For instance, compared with the iPhone, the iPad is more suitable for displaying larger size ads. As another example, the fact that a consumer owns an iPhone X will have consequences for consumption capacity. Such information may be utilised to target ads.

2. **Unique device ID.** Every mobile device is assigned an International Mobile Station Equipment Identity (IMEI) after it is produced. The IMEI, which is unique around the world and usually used to identify valid devices, can be used to support TMA. For instance, if several application developers partner with the same ad-network, they can share the specific user's information to create a more precise profile by locating the unique IMEI. In addition, even if an app is uninstalled from a given mobile device, the ad-network can still collect that user's data with other apps. If the app is installed again or a new app belonging to the ad-network is downloaded, they can continue sharing the user's data without missing anything by identifying the IMEI. Going further, sometimes the MAC address of a mobile device is also collected as the device ID. It is semi-permanent (unless the user deliberately modifies it).

3. **SIM card information.** The information stored inside the SIM card can also be obtained easily with APIs. A SIM card associated with a mobile phone contains its phone number and SIM series number. This information enables advertisers to establish a direct contact relationship with the user. In some regions, the phone number can also indicate the network provider and the network type of a mobile phone. For instance, a phone number starting with 139 in China indicates that its network type is 2G and the network provider

is the company *China Mobile*, while a phone number starting with 186 in the same country indicates that it is a 3G SIM card provided by *China Unicom*.

4. **System preferences.** Most smartphones require their users to set the system preferences when they start up the device for the first time. These preferences can provide some useful information to the advertisers: the system type and version can be used to determine which peripheral device ads should be delivered. Preferred language and region indicate the language that ads should use. Moreover, settings in system accessibility may indicate that the user has disabilities; this information can, in turn, be used to deliver targeted ads.

5. **Communication information.** Communication information includes contact information, call logs, messages, and emails. Ad-networks can use contacts and call logs to perform viral marketing. They can record users' consumption information by analysing the messages sent from their banks. For example, *Now cards* — a sub-app of the *Google* app — can automatically suggest a link to track packages by analysing a recent email confirming a purchase. With relevant permissions, which are often ignored by users, communication information is exposed to third-parties.

6. **Calendar events.** With users' authorizations, which are usually neglected by users and set to default values, apps can add, remove and change calendar events and send emails to guests without the user's knowledge. The capability of operating calendars makes marketing to specific people with specific themes much easier for those advertisers.

7. **Browser cookies.** Cookies on PCs enable ad-networks to track users' browsing activities across different web-sites. The cookies associated with the built-in browsers of mobile devices also record users' activity in much the same way. Together with the unique device ID, the tracking effectiveness of mobile devices can be more effective than on PCs.

8. **Installed apps.** As mentioned above, users' operations on apps carry some useful behavioural metadata. In fact, even without tracking the operations, advertisers can still create profiles of users' interests by analysing the number and categories of the apps that have already been installed on a mobile. The fact that a mobile device has one of *NBA Game Time* or *MyNBA2K18* installed is probably sufficient to indicate an interest in professional basketball.

9. **Basic personal information.** Most mobile systems provide a function for users to create their own contact card, which contains their names and contact information. This information can be collected easily by calling the relevant APIs.

2.3 Related Work

We now discuss previous work in the context of TMA and privacy-preserving approaches.

2.3.1 Previous research on TMA

In recent years, the proliferation of mobile devices has presented advertisers with new avenues: traditional advertising forms such as SMS or MMS advertising are gradually being replaced by in-app advertising. The marketing advantages of mobile devices, such as the capability to obtain users' locations, as well as information about and from installed applications and other contextual information from various sensors, have offered great benefits in delivering appropriately targeted advertisements. Meanwhile, mobile users' privacy concerns are increased. Studies of preserving privacy in TMA are also conducted.

In general, existing studies related to TMA can be classified into three categories: mechanisms for performing targeted advertising, mechanisms for preserving privacy, and mechanisms for compensating privacy leakage. The research described in this dissertation leverages techniques from each: our aim is to combine different mechanisms to provide benefits to both consumers and advertisers in the context of TMA.

Mechanisms for performing targeted advertising

Location-based advertising (LBA) is an advertising approach that aims to deliver relevant ads of services or goods that can be accessed near to where a mobile user is geographically located. With LBA, ad-networks can display more accurate ads to mobile users when they are most likely to make a purchase. MobiAd [14] is an approach for performing localised and personalised advertising. A unique feature of MobiAd is that the downloading, displaying and click-reporting of ads are all accomplished via a Delay Tolerant Networking (DTN) protocol, which provides anonymity for users when using wireless communication. MALCR [16] is another advertising approach based on a user's physical location. It makes use of two-level neural network learning to analyse users' profiles: the first neural network learns users' behaviours, and the second learns their interests. Apart from the particular technologies used in LBA, Zou *et al.* [21] conducted a randomised field experiment to analyse the LBA strategies crucial to advertisers.

Probabilistic reasoning advertising differs from LBA in that it aims to predict the best ads for a specific user by not only analysing the user's current location, but also by exploiting the user's behavioural patterns. For example, delivering an ad of a nearby pizza restaurant to a user by simply taking into account the user's current location and food preferences might not be appropriate, especially if the user has just visited another pizza restaurant. To this end, AdNext [22] provides a visit-pattern-aware mobile advertising approach. By analysing sequential visit patterns of users, AdNext can predict places that they are likely to visit in a city and deliver them relevant ads. The key mechanism of AdNext is a probabilistic prediction model that helps to make the prediction on the basis of users' visit histories.

User-profile-based advertising aims to exploit users' information to construct stereotypes. A user profile might indicate the user's age, gender, income, interests, pattern of daily routine, etc. MALCR [16] filters ads based on some relatively simple user attributes. Another approach developed by Bilchev and Marston [23] also delivers personalised advertising based on user profiles, with a key feature

being that the approach is sensitive to the fact that a single user's profile might be distributed across devices and accounts.

Mechanisms for preserving privacy

Hybrid personalisation mechanisms are applied in a number of contributions to protect users' personal information while serving ads. The hybrid personalisation mechanism allows users to pre-download several ads from ad-networks with a generalised context, and then select the best one on the client on the basis of accurate user information. Adnostic [3], Privad [24] and MobiAd [14] are examples of this approach.

Advertising frauds defence is a significant mechanism for a reliable TMA system. In addition to user-targeting approaches, the ability to defend against view-fraud or click-fraud attack is also a key factor for a usable advertisement-selection mechanism. Many approaches, including AdAttester [25], DECAF [26], MadFraud [27] and Bluff ads [28], have made efforts to detect and prevent advertising fraud attacks.

Data aggregation mechanisms are also widely applied in TMA as a privacy-preserving data collection method. The data aggregation approach ensures that only statistics over the data contributed by multiple mobile users can be collected by advertisers, while individuals' personal information is preserved in their own devices. As an example, Zhang *et al.* [29] proposed such an approach that makes use of information hiding and homomorphic encryption to guarantee the data privacy of mobile users.

Mechanisms for compensating privacy leakage

Valuations of users' privacy are treated as a part of the users' utility by Nissim *et al.* [30]. This view suggests that users should be able to trade their personal data for benefits according to their own wishes. To this end, several mechanisms are proposed in TMA for performing targeted advertising by compensating for users' privacy leakage. For example, in the targeted advertising framework proposed by Wang *et al.* [31], the ad broker (ad-network) receives money from advertisers and pays compensation to users for clicking related ads. Thus, users can determine their

clicking behaviours based on the amount of compensation and level of sensitivity of their exposed information. Although users are aware of their privacy leakage, the compensation mechanism encourages them to get involved in the privacy-trading process.

Self-selecting advertising also helps to reduce users' hostility towards advertisers, and helps to increase response rates. Recognisable benefits such as special discounts or details of products that consumers are interested in often works well in attracting users to accept mobile ads. In such cases, some users might subscribe to specific brands to obtain ads of new products as soon as possible or explicitly request ads from ad-networks to enjoy discount coupons. For instance, Shopkick[1] is an app that can provide different rewards such as movie tickets, gift cards, or discount coupons to users. To obtain the rewards, users need to explicitly pick products with the app that they are interested in.

2.3.2 Previous research on privacy-preservation approaches

The related technical approaches that try to preserve users' personal information can be initially classified in five research clusters: client-based solutions, manual intervention solutions, context obfuscation solutions, sandbox solutions, and other solutions. We discuss them in turn.

Client-based solutions

Most current advertisers apply targeted advertising by collecting and analysing consumers' personal data at their servers. To address privacy concerns in this situation, several prior contributions have provided approaches that keep personal data at the client device and perform targeting locally. We characterise this kind of approach as client-based. Hybrid personalisation mechanisms (as discussed above) are applied in most client-based approaches.

Adnostic [3] allows PC users to provide only coarse-grained information about themselves while detailed personal profiles are used to select relevant ads and kept

[1] `http://www.shopkick.com/`

locally. Its prototype is implemented as a Firefox extension with two modules: the user-profiling module builds a list of user interests inside the browser and the ad-rendering module selects ads to be display locally. The other feature of Adnostic is that its billing system can report which advertisement was viewed without exposing this to the ad-network, so that the ad-networks can charge advertisers for impressions in a privacy-preserving way. This feature is supported by using zero-knowledge proofs and homomorphic encryption. The number of views for each advertisement are sent to a trusted third-party to decrypt them first, and then returned to the ad-network. However, there exists a significant limitation of Adnostic: it can work only with the cooperation of ad-networks. Ad-networks need to modify the way they serve ads in order to support Adnostic.

The Privad project [24, 32] has a similar goal to Adnostic. Privad contains four entities: client, dealer, monitor and broker. In this model the system downloads all potential ads from the broker, and selects the appropriate ads on the client. The dealer works as an anonymous proxy between the client and the broker, so that the broker cannot identify the client. Moreover, communications between the broker and the client are encrypted with a public key mechanism, therefore even the dealer cannot fetch the information about which ads have been downloaded. The monitor is used to ensure that the client cannot send information to the broker through a covert channel. In addition, the Privad project team proposed a manual intervention mechanism, which uses subscription-based prefetching to enable users to access the relevant ads whenever a proxy is available. To realise that, users need to manually subscribe to the categories of ads that they are interested in before they start browsing. Unlike Adnostic, Privad does not trust ad-networks at all and it anonymises all of the information sent by the client. Consequently, the anonymising operations during its function would impact performance and consume more resources to detect click-fraud. Since the pay-per-click model is so popular, the approach adopted by Privad is unlikely to find favour with advertisers.

MobiAd [14, 33] is a private advertising system for mobile platforms. Ads are selected locally from the pool of ads, and only those ads related to users' interests

will be downloaded, just as Adnostic has done on PC browsers. The information for user analysis such as ad views and clicks are encrypted and sent to the ad-network via intermittent Wi-Fi hotspots and other phones. Therefore the advertiser can only receive aggregate information and cannot distinguish between users. MobiAd has the same drawback as Adnostic: the system needs support from the ad-networks.

The above systems are all client-based solutions. As we can see, the basic principle of these solutions can be described as follows: the client provides only coarse-grained information to the server, then the server chooses a set of potential ads (in some solutions, the full set of ads) for the client to download, and finally the client runs the ad-selection algorithm locally and chooses the most suitable ads from the downloaded set to display.

Obviously, the biggest advantage of client-based solutions is that most of the personal information is kept locally inside the client, therefore consumers' privacy is preserved. However, it is hard to assess the number of ads that should be put into the potential set and the effectiveness of the ads that are finally displayed. Some of the solutions choose to cooperate with ad-networks, therefore they can be deployed smoothly and share the most efficient ad-selection algorithm with ad-networks. But it is not easy to persuade the ad companies to modify their existing infrastructure and give up consumers' personal data, which they can easily obtain without their cooperation. The other solutions can be deployed more flexibly without the support of ad-networks; however, the accuracy of selected ads and their performance cannot be guaranteed easily.

In addition, it is unavoidable that such solutions would introduce certain impacts of performance on user experience. However, with appropriate configurations, the impact might degrade to a tolerable level. For example, Adnostic minimises the impact by loading multiple ads after the enclosing page is rendered and by limiting the size of the list of ads that needed to be pre-downloaded. As another example, Privad states that, by eliminating network round-trips during the process of rendering webpages, users' web browsing experience might be improved, as the page load time would be much faster without unnecessary redirect.

Solutions based on manual intervention

There is an important variable to which we need pay close attention in evaluating each kind of solution: the level of privacy exposure. We have to decide how much information about the consumers' contexts should be shared with the ad-networks. Even in the client-based solutions we need to answer some questions about which kind of 'coarse-grained information' the client shall provide to the server for the potential ad-selection. For example, does the information only contain the consumer's broad interest category, or should the location and gender also be included? Does one configuration apply to every ad-network or do different ad-networks have different configurations? To determine privacy variables, some prior contributions have proposed manual intervention solutions.

RePriv [34] is a policy architecture that allows users to control the extent to which their data is shared with the ad-network. In this model, the browsers can mine users' personalised information locally and create related interest profiles. Online service suppliers can register their own miners to extract the additional information from users' operations. Each time they ask for the user's data, an explicit prompt asks the user whether specific data from the interest profile can be submitted to the supplier. Therefore, users have complete control over their personal information, as long as they can tolerate the regular prompts.

Another approach suggested by Hardt and Nath [35] proposes a framework targeted at mobile advertising with a manual intervention mechanism. Users can decide how much of their contextual information they are willing to share with the ad-network. Then the ad-network selects a list of ads for the client by computing the limited information submitted by the client. Finally the client can pick the most relevant one from the list of ads based on all information about their private context.

Lockr [36] is an access control system that separates the data management of OSNs (Online Social Networks) from their other content-sharing functionality. Therefore the users can decide which service supplier could store their own social information, and which third-parties should have access to it. In Lockr, the social relationship between two users is signed digitally. The recipient of it can then

provide the relationship to the social network as proof to obtain the issuer's social data while the relationship cannot be reused by the service supplier. However, the management of relationships might become quite inconvenient after the user has established a significant number of relationships with others.

The solution Virtual Individual Servers (VISs) [37], which could be deployed as privacy-preserving proxies for mobile devices, offers a choice for users who want to enjoy online services on mobile devices without giving up control of their data. A VIS is a virtual machine running in a utility computing infrastructure. It is resistant to large-scale privacy breaches because each VIS runs in its own administrative domain. A VIS gives its owner thorough control over the software and policies that determine what data is shared with whom. The manual intervention mechanism of Lockr and VISs could also be applied in targeted advertising for users to adjust privacy-sharing levels for different ad-networks on their own initiative.

The above projects are some examples of manual intervention solutions. The basic principle involves giving consumers explicit control over the sharing of their personal information. Some of the solutions maintain configuration files for different service suppliers, and some solutions give all control back to the consumers, requiring the consumers to make decisions every time their personal information is required. The main strength of manual intervention solutions is that consumers have complete control over their personal data. In addition, the manual intervention mechanism makes the process of sharing data much more accurate and flexible. However, there are significant drawbacks in these solutions: for the solutions with configuration files, the management of configuration might become quite inconvenient with the increased involvement of the ad-networks; for the solutions that need consumers' intervention each time, the constant prompts might become intolerable. To mitigate these problems, some solutions use typical configurations to simplify the management, while others make use of different strategies for their systems to learn user preferences and answer the prompts automatically.

Context obfuscation solutions

Unlike the solutions already discussed, there are some approaches that share no pieces of personal data with online service suppliers. Some of them encrypt the real personal data into a fake, but legitimate, data form and send them out; others submit the real behaviour together with several fake behaviour to obfuscate from the server. We describe this kind of approach as context obfuscation solutions.

In NOYB (short for none of your business) [38] researchers encrypted user data, and encoded the cipher-text to look like legitimate data on which the online service can operate. This means that the online service can not discover those users who use the fake data to protect their privacy. In addition, their original data can only be decoded and decrypted by the authorised users. NOYB was implemented as a browser plugin for the Firefox web browser. With this plugin, the information *(Alice, Female, 24)* can be faked to *(Bob, Male, 25)*. Only the authorised friends of Alice can obtain her real information. Therefore, the online service suppliers cannot obtain the user's real personal data while the communication between different users would not be affected. However, to use NOYB, a comprehensive dictionary, which is used for mapping the original and fake information, is needed. In addition, since Alice was treated as a male by the online services, she cannot get the advertising service targeted to females anymore. This means that some of the consumers' benefits are lost with the application of NOYB.

However, NOYB still provides a good starting point for us. One feature of NOYB is contextual integrity, which is described thus: "In such a framework, pieces of a user's information are scattered but public; it is the inability of an adversary to combine these pieces that defines privacy" [38]. For instance, the online service knows that there exists some user named Alice, some user is female, and some user is 24 years old. But the online service cannot confirm that there is a female user whose name is Alice and aged 24. Another feature of NOYB is incrementally deployable. NOYB can be deployed incrementally by small groups of users without any cooperation from the online service.

TrackMeNot [39] is a tool also implemented as a Firefox extension that attempts to hide user queries to search engines by submitting random fake queries periodically. This method can also be applicable for defeating behavioural tracking: when a user clicks on an ad or visits a page displaying that ad, the system will 'click' or 'view' random ads or pages during the same session, therefore the ad-network cannot identify the user's real interest. However, if TrackMeNot has been deployed on a huge scale, the enormous fake queries would distinctly degrade the performance of search engines or ad-networks.

On the mobile platforms, the ad-networks or related analysts could develop users' interest profiles by analysing the installed apps on their mobiles and the use of different apps with specific frequency. For example, the installation and frequent uses of sports and business apps with frequent uses of them could indicate that the user has an interest in those areas. ProfileGuard [40] is an app-based obfuscation mechanism that works on mobile platforms that prevents third-party apps from inferring user interests by analysing installed apps. One implementation of this is in terms of an Android app. Users can run a customised ProfileGuard app and select their private interest category to protect the specific aspects of their profile. This app can analyse the information of those apps which have already been installed on the mobile device and suggest the potential threat with candidate-obfuscating apps to the user. The user could then select obfuscating apps to install and run so that the ad-networks cannot determine the user's real interests. The list of obfuscating apps is generated under an obfuscation strategy. The first such strategy is based on the most similar apps from the user's non-private app category, while another is customised to the user's profile interests. ProfileGuard helps users retain their privacy, however the obfuscation means that users cannot benefit from accurate personalisation-targeted online services.

MockDroid [41], a modified version of the Android system, allows users to replace sensitive personal information with fake data. For example, if a third-party app asks for the user's location information, MockDroid could provide the fake data by reporting a failure message such as "lack of a GPS signal". However, as a

modified Android system, MockDroid is relatively complicated and difficult for users without the relevant knowledge and skills. In addition, it has the same drawback as ProfileGuard in that users cannot get information that they are interested in because of the fake information.

The above systems are all context obfuscation solutions, and can be classified in two clusters: the first one transfers the real data to fake data; the second one submits real data together with fake data. The consistent principle is the obfuscation of the user's privacy context so that the online service suppliers cannot obtain the user's real personal information. Since all data and operations involved in these solutions are legitimate to the online services suppliers, the deployment of these approaches does not rely on the support of the suppliers. With these solutions, most of the consumers' personal data would not be exposed to the suppliers. However, users also lose their benefits of obtaining personalisation targeted services due to the obfuscation mechanism. Furthermore, the enormous fake operations would degrade the performance of the network.

Sandbox solutions

In computer security, a sandbox mechanism is often used to execute untrusted programs from unverified suppliers. There are some contributions that import the concept of a sandbox to preserve personal data.

AdJail [42] is a model that protects users from untrusted ads with policies and sandboxing mechanisms. In this model, ads will be isolated from the original page to a shadow page. Users and ad-networks interact with each other via the shadow pages under its specified policy. Therefore it successfully isolates malicious ads from the Document Object Model (DOM) of the hosting page; however, it cannot completely protect security-sensitive APIs such as local storage and XHR due to its isolation feature. In this regard, AdSentry [43] is quite similar to AdJail. AdSentry executes untrusted JavaScript ads in a shadow JavaScript engine sandbox on the client. Under its access control policy, the client accesses the sandbox instead of accessing the DOM and sensitive functions directly.

TaintDroid [44] is a popular information-flow tracking system for the Android platform. It monitors untrusted apps and can be used to analyse how those apps access and manipulate users' sensitive personal information. DroidBox [45] is an Android application sandbox that uses TaintDroid to detect the leaking of users' personal information. It can provide a timeline view of the monitored app's behaviour for users to identify malicious apps. In addition, there are also many other Android sandbox architectures such as DroidScope [46] and Mobile-Sandbox [47].

Sandbox solutions might work in two ways in the context of TMA: they could isolate malicious ads to prevent the disclosure of sensitive data and they could work as a monitor to analyse the privacy-unfriendly behaviour committed by third-party apps. These kinds of approaches generally have limitations such as not being able to detect particular attacks or not being able to monitor all possible leaks. However, a sandbox mechanism is still a potential method to improve the effectiveness of preserving privacy under certain circumstances.

Other solutions

Besides the approaches discussed above, there are several other solutions that could also be applied to preserve privacy in TMA. We describe them briefly below.

The developers of FindU [48] proposed privacy-preserving personal profile matching schemes for mobile social networks. In FindU, an initiating user can find from a group of users the one whose profile best matches his or hers, to limit the risk of privacy exposure. This scheme, with some modifications, can also be applicable for privacy-preserving TMA, for example, to create several virtual characters to apply for the targeted ads, then match users with the characters to obtain the related ads.

Some researchers have proposed designing online network applications based on a peer-to-peer architecture in order to avoid centralised control over a user's data. Among those peer-to-peer based solutions, Safebook [49] stores each user's data at the nodes of that user's trusted friends, and with additional feature to assure the intractability of the communications during look-up and data retrieval operations.

Cutillo *et al.* [50] propose another architecture also based on P2P to solve privacy issues in OSN. Their model consists of three components: mappings that are used to provide the basic distributed structures that contain the users' personal data of different OSNs; a peer-to-peer substrate that provides the global access to a user's profile; and a trusted identification service that provides authentication. Thus the decentralized OSNs should distribute users' personal data across different domains and the chance of privacy breaches could be reduced.

The simple reputation scheme of the iClouds Project [51] provides a method to protect privacy properties while sharing information. It uses public key functions as a user's pseudonym. All messages between two parties contain the pseudonym. It could be applied to TMA via submitting a pseudonym instead of a traceable user identity to the ad-networks, meaning that the ad-networks cannot identify which user views which ad.

Juels [52] proposed a personalised advertising deliver protocol based on Private Information Retrieval (PIR) and mix networks. This scheme focuses on the private distribution of ads and does not consider other aspects such as view-and-click reporting or auctions. In this protocol, the server does not peremptorily deliver the ads; instead, the client requests relevant ads which should be delivered later when the client visits the page again. By relaxing the conventional PIR mode, the scheme is able to achieve considerable practical improvements in terms of both communication and computational complexity. However, in terms of efficiency, this protocol demonstrates low-performance, because an ad of a given page needs to be prefetched in order to be displayed later when a user visits the page again.

ObliviAd [53] is another project based on the PIR protocol that provides targeting ads without exposing consumers' information. The special aspect of this model is that it requires specialised hardware located at the ad-network. This hardware creates a black box using the secure co-processor and oblivious RAM. The black box is responsible for distributing ads to clients, receiving click reports from clients, and recording view data for each ad without identifying which client has viewed it. After this process, it releases only the view records to the ad-network in large batches

so that the server cannot determine which client clicked which ad. There is an obvious special drawback of ObliviAd: it can only be deployed on those ad-networks which are compatible with the specialised hardware.

2.4 Summary

In this chapter we have provided the background information about the current TMA ecosystem. We have discussed the characteristics of mobile advertising, as well as the different kinds of personal data stored in mobile devices. We have reviewed previous research in the context of TMA. In the following chapter we will report upon the results of a qualitative study to discuss users' perspectives on privacy and discuss the balance that needs to be struck between privacy and utility in this emerging area.

3

Perspectives on privacy

Contents

3.1 The trade-off between privacy and utility 32
3.1.1 The unavoidable trade-off 32
3.1.2 The underlying free app ecosystem 33
3.1.3 Attitude towards the trade-off 34
3.1.4 The privacy paradox 36
3.1.5 The authenticity of collected information and the lack of privacy awareness 37
3.1.6 A real-life example 38
3.2 Human factors contributing to the trade-off: a user study 43
3.2.1 Overview of the user study 43
3.2.2 Study approaches and participants 45
3.2.3 Scenarios and data collection 46
3.2.4 Results and discussion 49
3.2.5 Lessons learned from the user study 59
3.3 Requirements and threats 62
3.3.1 Scope 62
3.3.2 High level requirements 63
3.3.3 Threat models 65
3.4 Summary 66

This chapter discusses people's perspectives on privacy issues pertaining to mobile services, and provides high level requirements to help address our research problem. Section 3.1 gives an overview of people's attitudes towards the trade-off

between privacy and utility in mobile services. Section 3.2 then reports the results of a user study that investigates the factors that influence the decision-making process pertaining to the trade-off. Based on the results, Section 3.3 discusses the potential improvements and provides high level requirements for a privacy-preserving system that has the potential to help users in their decision-making. Then in Section 3.4 we give the summary of this chapter.

3.1 The trade-off between privacy and utility

3.1.1 The unavoidable trade-off

Today's smartphone is not only a communication device; it is also a mini-computer that supports people's daily lives. The capacity of dynamic computing, and the integration of various sensors, together with the explosive growth of apps, make it possible to provide numerous utilities — which used to require expensive hardware — via relatively cheap mobile services.

However, while the widespread use of mobile services offers a variety of benefits to mobile users, it also raises serious privacy concerns. According to Appthority's Winter 2014 App Reputation Report [54], 95% of the top 200 free iOS and Android apps and 80% of the top 200 paid apps exhibit at least one behaviour that might be considered a privacy risk. In particular, of the top 200 free apps, 70% allow location tracking, 56% identify users, and 53% share personal data with ad networks. These findings suggest that almost every time people attempt to use mobile services they are making a decision to exchange their privacy for benefits.

For example, people need to provide their location to obtain a real-time weather forecast or expose their interests to acquire accurate recommendation of goods or activities. Indeed, the economic model behind most free mobile services is based on such trade-offs. For instance, when using mobile services provided by Facebook, Google and Twitter, users do not directly pay the service providers money to download or use their mobile services. However, the mobile platforms collect a vast amount of users' personal data by analysing users' in-app behaviour. The personal data will then be used to sell highly targeted ads.

People's attitudes towards such privacy issues vary significantly. Some users would be prepared to abandon the benefits provided by mobile services to protect their personal data on their mobile devices. According to a survey of 2,000 Americans conducted by Pew Research Center [55], 54% of mobile users have decided to not install an app when they discovered how much personal information they would need to share in order to use it and 30% of mobile users uninstalled an app after discovering it was collecting personal information that they did not wish to share. Some others would enjoy the benefits without considering the potential privacy risks. A field experiment [56] showed that about 93% of participants are willing to provide personal data about their date of birth and monthly income for a 1 Euro discount for purchasing DVDs.

3.1.2 The underlying free app ecosystem

Statista [57] in 2016 reported upon the average daily time spent online via mobile by users worldwide. As of the reported year, Generation Y (defined as individuals who reached adulthood around the turn of the 21st century) internet users spent an average of 185 minutes on mobile internet services every day. The average daily time spent by Generation X (defined as the demographic cohort following the baby boomers) users reached 110 minutes. The increasing mobile phone internet penetration indicates the important fact that mobile devices are becoming indispensable tools in people's daily lives.

As two of the biggest mobile operating systems, Android and iOS now have presented more than 2 million apps respectively in their official app stores [58]. Both Google Play for Android and Apple App Store for iOS offer apps that cover different categories from games to business, and from education to shopping. The diversity and large number of apps make it possible for people to enjoy the convenience provided by mobile services for a wide array of purposes.

Compared to the huge benefits received by users, the cost of these mobile services is, on the face of it, extremely cheap. Users pay the bill with their personal information instead of cash. According to AppBrain [59], by the end of 2017 only

about 222,600 of the total 3.5 million apps on Google Play were paid apps, which means more than 3.3 million apps are all free to download. In addition, only about 275,700 of these free apps offer in-app purchases — most of the rest are totally free to use and rely on in-app ads. The free app business model unavoidably results in the collecting of a vast amount of users' personal data, as these apps need individuals' data to make the in-app ads highly targeted.

Many of these apps also share the collected data with third-parties. As reported by a study conducted by Zang *et al.* [60], in the 110 most popular free Android and iOS apps they tested, about 73% of Android apps and 47% of iOS apps shared personal information with third-parties. The shared data ranges from personally identifiable information such as name, gender and email address, to in-app behaviour data like job searches, chats posts and medical information. In addition, users' current locations collected by built-in GPS sensor of mobiles are also frequently collected and widely shared.

The highly involved mobile services and the free app ecosystem regularly lead mobile users into situations requiring them to make decisions. Almost every time they attempt to use a mobile service, they are making a decision to exchange their privacy for benefits.

3.1.3 Attitude towards the trade-off

It is important to note here that not all data sharing equally bad. On the one hand, the data-sharing approaches and related targeting technologies in use today have made a large growth of revenue for retailers and advertisers, which, in turn, drove the world economy. On the other hand, many people believe that they should have the right to share their own data in their own best interest without regulative intervention [19]. However, unlawful data sharing might do harm to consumers physically or mentally, might make adverse impact on business profit, and might reduce consumer confidence and raise privacy concerns. Hence, related laws —

with the EU General Data Protection Regulation (GDPR)[1] as a representative — have been proposed to constrain such activity

From the perspective of consumers, people's attitudes toward the trade-off between privacy and utility vary significantly. The results of a survey conducted by Marketing Land [61] suggest that close to half (46%) of respondents indicated either a neutral or unconcerned reaction to being retargeted by retailers, while the rest have concerns about the potential privacy risks.

Individuals behave inconsistently: not only do different people's attitude vary, but the same user may choose different strategies to handle the trade-offs when using different mobile service as various factors might be involved in their decision-making process. For instance, when facing the trade-off between receiving the benefits of a mobile coupon and giving up personal information, factors such as perceived value, personal innovativeness and coupon proneness could positively affect the user's acceptance of related apps, whereas perceived fees and perceived privacy risk might have negative influence [62].

When using location-based mobile services, factors like compensation, industry self-regulation and government regulation [63], or personality traits such as agreeableness, extraversion, emotional stability, openness to experience and conscientiousness [64] can all affect a user's decision.

The inconsistency between people's privacy attitudes and their final decisions makes the situation complicated to study. Deciding to use an app does not necessarily indicate a lack of concern about privacy. Users might take a number of steps such as clearing their browsing and searching history or turning off the location-tracking feature on their mobile to protect their personal information before enjoying mobile services. On the other hand, users who claim they have a serious concern about privacy risks might also share their personal information willingly for certain benefits. The latter phenomenon — people's observed behaviour being inconsistent with their reported attitudes — is usually termed the *privacy paradox* [65].

[1] https://www.eugdpr.org/eugdpr.org.html

3.1.4 The privacy paradox

Many authors, including Awad and Krishnan [66], Barnes [65], Norberg *et al.* [67], and Utz and Krämer [68], have given consideration to the term *privacy paradox*, which recognises the inconsistency between people's reported attitudes and their observed behaviour when interacting with online services. As social exchange theory [69] explains, if the exchange is perceived to be beneficial, then the individual is likely to enter into an exchange relationship. With regard to TMA, users exchange the benefits of useful mobile online ads with their privacy. Sometimes consumers' unwillingness to pay for privacy can be rationalised with relatively small rewards. For example, the field experiment discussed earlier [56] involving two stores showed that participants are willing to provide personal information such as their monthly income and date of birth for only a 1 euro discount. Surprisingly, and inconsistent with their behaviour, about 82% of the participants reported a decreased willingness to provide their income information. In fact, participants predominantly chose to shop in the store with the lower price but the requirement for more sensitive data. Moreover, the experiment also shows that even though prices, shopping time and delivery time were equal at both stores, the store that captured less personal information failed to attract more customers.

The great protections claimed by privacy notices also contribute to users' relatively lax privacy attitudes, in which case privacy violations can be caused by users' over-reliance on privacy notices. Martin [70] suggests that users have been found to assume the protections offered by privacy notices can meet their privacy expectations even when some stated notices differ considerably from the actual ones. Further, privacy notices are likely to be ignored due to the long text and obscure language [71, 72]. As a result, consumers may adopt a blind reliance on firms' business ethics.

On the other hand, research undertaken by Tucker [73, 74] suggests that the opposite is true: that consumers still pay attention to their personal information while using online services. An experiment undertaken by Tucker [74] showed that a Facebook page was twice as effective at attracting users after the Facebook policy

gave users more control over their own personal information and made the privacy settings easier to use. The experiment could imply that, by giving consumers explicit control over how their data will be used, the advertisers may be able to alleviate some of the hostile awareness caused by violating users' privacy while performing OTA.

It is understandable that different kinds of people may have different privacy concerns. If one were to take 'age difference' as an example classifier, it is commonly thought that adults are more likely to have a stronger awareness of privacy risks than teenagers. For example, a study conducted in 2012 by Madden *et al.* [75] showed that 81% of parents of online teenagers said they were concerned about how much information advertisers can learn about their child's online behaviour, while the related figure for teenagers (with respect to their own privacy) was only about 9% [76]. A further study by Madden [77] in 2014 also demonstrated that 80% of adults who use online social networks say they are concerned about third-parties like advertisers or businesses accessing the data they share on online social networks. Another possible classifier is 'occupation difference': some occupations, for example, those serving in the military or working in healthcare, have rules or guidelines about how much personal information the person can share online.

3.1.5 The authenticity of collected information and the lack of privacy awareness

The ways in which online advertisers collect users' personal information can be classified into two categories according to different initiators: the first is when advertisers capture and analyse users' behaviour or collect context variables from various sensors to obtain personal information; the second is when users submit personal information voluntarily. Most data collected in the first way can be presumed to be real and objective; information obtained in the second way is less likely to be authentic, as users could offer fake data. Several authors have proposed surveys to determine how truthful information returned is. For example, in 2005, Marwick [78] randomly selected 30 user profiles from three different social networking sites (Friendster, MySpace and Orkut) and analysed the pictures, text

and testimonials on each. The result showed that the great majority of user profiles (72 out of 90 — 80%) were presumed to be authentic. Although the sample was relatively small, the study did demonstrate some characteristics of the users of online services. In 2008, Tufekci [79] used a much larger sample (n = 704) of college students to discuss the information disclosure in online social networks. A significant majority (94.9%) of the sample said that they used their real name on Facebook, although there is no obligation to submit their real information. A report [76] in 2012 also demonstrates the phenomenon. Researchers asked 802 teenaged social media users about ten different categories of personal information that they might post on the profile they use most often and found that: 92% of them post their real name; 91% of them post a photo of themselves; 84% of them post their real interests, such as movies, music or books they like; 82% of them post their birth date; and 71% of them post the city where they live. These studies indicate that the great majority of online users submit real personal information to the service suppliers voluntarily, which provides detailed data for advertisers to apply TMA. The lack of awareness of privacy risks might account for this phenomenon. For instance, in [76] it is reported that teenaged users do not express a high level of concern about third-party access to their data: only 9% say they are 'very concerned'.

3.1.6 A real-life example

Having discussed the unavoidability of the trade-off, the underlying free app ecosystem and users' various attitudes, we now present a real-life example leveraging the Android system to better introduce these factors.

Android apps must declare a list of exact permissions they require to execute — for example, whether the app needs to read the user's contacts, obtain the user's location information, or modify the user's calendar; this list shows what kinds of access the application has to the user's personal information. These permissions must be declared in the manifest file of the app, and are presented to the user in the first stage of the installation procedure. For example, upon the installation of Facebook (version 27.0.0.25.15), the user is presented with a list showing that

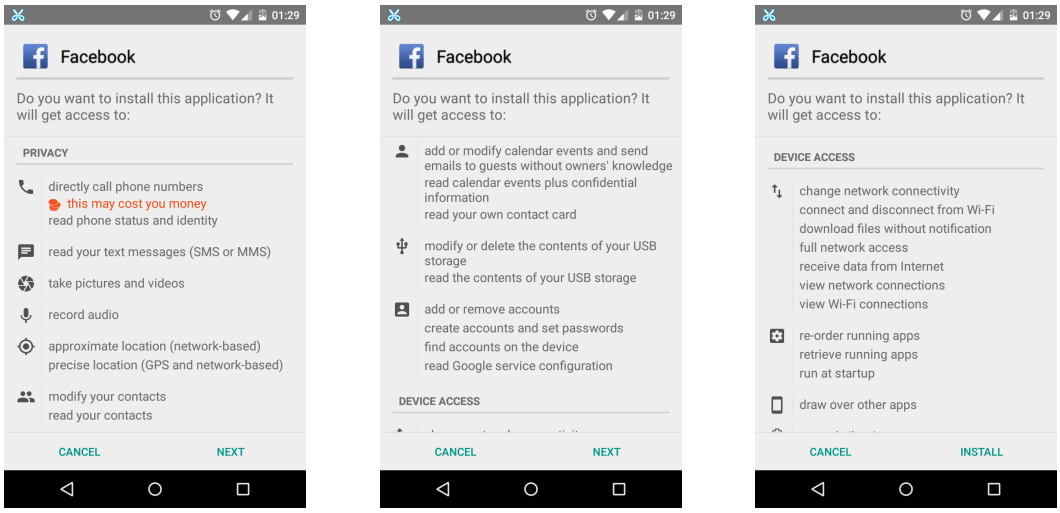

Figure 3.1: Facebook permissions list

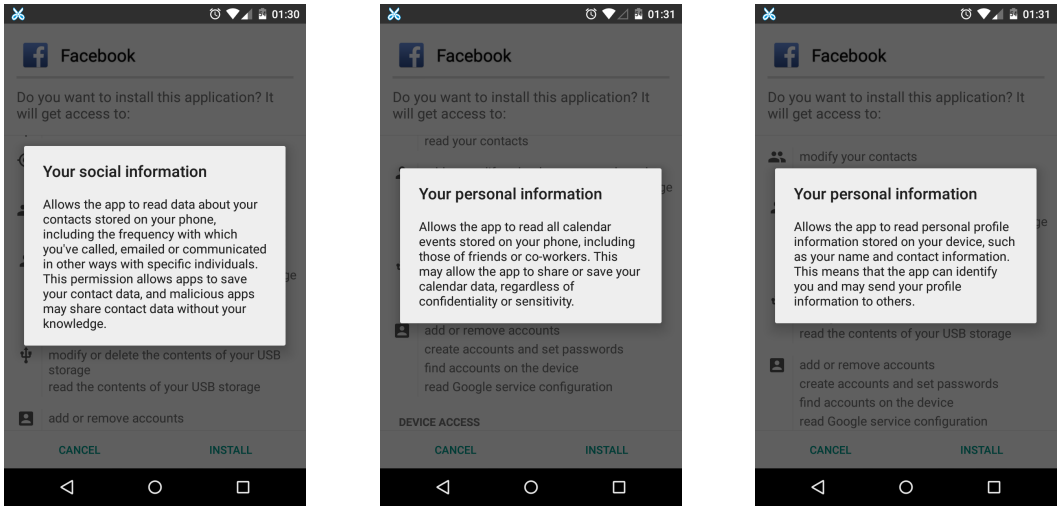

Figure 3.2: Permission details

the application requires 39 permissions, 18 of which are related to privacy and security (see Figure 3.1). When clicking on an item of that list, a more detailed description is given (see Figure 3.2). Having been presented with this information, the user can then decide whether to install the application.

The approach enables users to determine whether an app is likely to demonstrate unwanted behaviour. Three weather apps are shown in Figure 3.3. The 3D Digital Weather Clock app does not require any special permissions; the BBC Weather app asks for the user's location, which is used to provide localised forecasts. The AccuWeather app, however, requires the user's identity, the Device ID and call

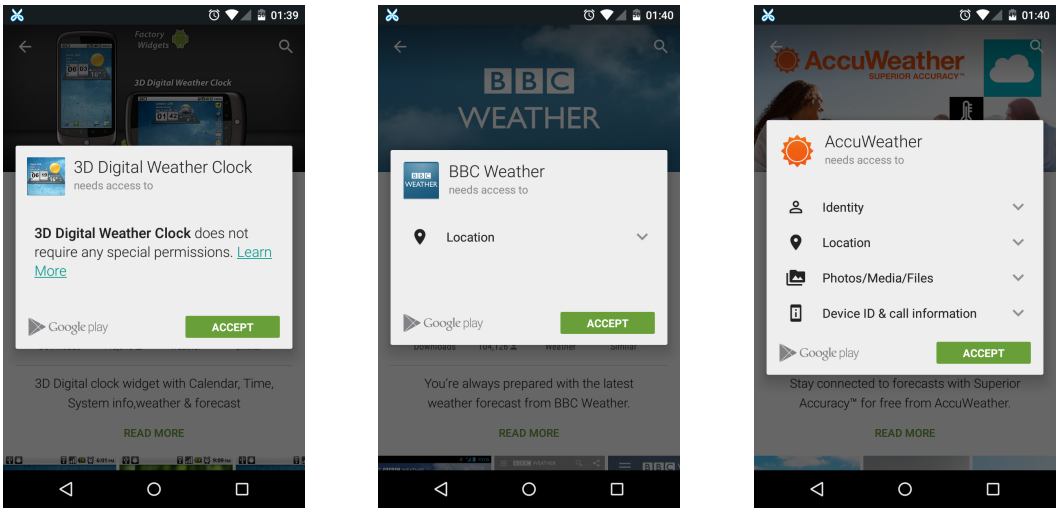

Figure 3.3: Permission lists of three weather apps

information. None of this is relevant for weather prediction, but is useful for, for example, targeted advertising. Based on this information, users can make informed decisions as to which of the apps they might wish to install. It is worth noting, however, that the reality does not match our expectations. In fact, at the end of January 2015, there were approximately 65,000 5-star comments for the 3D Digital Weather Clock, but more than 599,000 5-star comments for AccuWeather, which requires the most permissions of the three.[2] AccuWeather does feature targeted mobile ads (see Figure 3.4, where an advertisement for a local Asda supermarket is presented to an Oxford-based user), and its popularity may very well be due to its functionality: it provides a user-friendly interface and powerful extra functions. Nevertheless, many users have submitted negative ratings on the basis of privacy and targeted ads: some users never clicked on the ads; some others uninstalled the app immediately after they realised that it collects their personal information for advertising purposes.

To conclude, the permissions mechanism of the Android system is powerful, yet there are (at least) two significant problems for users that need to be addressed. First, many users with a lack of privacy awareness install apps without considering the permissions: a study [80] shows that only 17% of 308 Android users paid attention

[2]The first version of the 3D Digital Weather Clock has been available since June 2010, while the first version of AccuWeather has been available since January 2010.

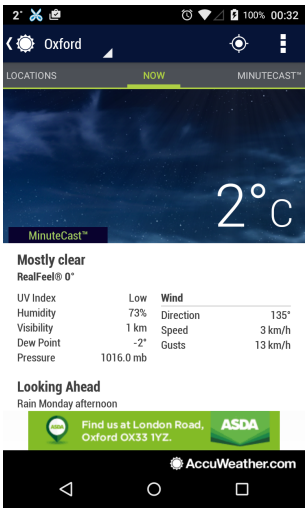

Figure 3.4: An app with targeted advertisements

to permission information during the installation, and only 3% of 25 participants could correctly answer three questions pertaining to permission comprehension. Second, even if a user has carefully analysed the permissions, the model provides a 'take-it-or-leave-it' offer: the user must permanently grant all permissions in order to install the app. It may be argued that, to give users more flexible control over their personal information, a complementary mechanism would be beneficial. To this end, proposed permissions extensions for Android include PermissionTracker [81], Flow Permissions [82] and Apex [83].

An improved permissions mechanism was used in Android 6.0 Marshmallow, which allows users to enable or disable permissions of apps after the installation in the new system. However, there are still two outstanding issues.

First, the new version of Android is not expected to support all existing Android devices. The detailed Android distribution numbers can be found in Table 3.1, which was updated by Google on August 8, 2017 [84]. Although Android 6.0 was officially released on October 5, 2015, more than half of the Android devices could not be updated to 6.0 or higher versions of the system by August 8, 2017. It is reasonable to believe that the Android 6.0 system or later versions can only be run on newer mobile devices. Therefore, a third-party permission framework is still necessary for a large number of devices that cannot be updated to Android 6.0.

Table 3.1: Android distribution numbers for August 8, 2017

Version	Codename	API	Distribution
2.3.3-2.3.7	Gingerbread	10	0.7%
4.0.3-4.0.4	Ice Cream Sandwich	15	0.7%
4.1.x	Jelly Bean	16	2.7%
4.2.x	Jelly Bean	17	3.8%
4.3	Jelly Bean	18	1.1%
4.4	KitKat	19	16.0%
5.0	Lollipop	21	7.4%
5.1	Lollipop	22	21.8%
6.0	Marshmallow	23	32.3%
7.0	Nougat	24	12.3%
7.1	Nougat	25	1.4%

Second, the new feature of modifying permissions after installation can only be supported satisfactorily with the new Android M SDK. Although the operation of denying permission can also be performed on apps designed for older versions of Android, it may cause the modified apps to no longer function as intended, as prompted in Android 6.0. By contrast, third-party solutions, such as returning fake or default values in the same format as the real ones, rather than simply denying permissions, can help to avoid this situation. Therefore, users will still need third-party frameworks to obtain more flexible control over the permissions of apps developed with older Android SDKs.

This real-life example suggests the need of new approaches that might help to reconcile the relationship between stakeholders of the trade-off. To explore the desired adoption of such privacy-preserving approaches, the results of presenting the core ideas of our proposed approach to participants in our user study are summarised in the following section

3.2 Human factors contributing to the trade-off: a user study

3.2.1 Overview of the user study

Online users' privacy concerns may vary depending on the situation [85]. In terms of mobile services, factors like perceived convenience, perceived financial savings, expected degree of information disclosure, and potential privacy risks all play important roles and lead people to different decisions. However, many questions still remain unanswered in this context. What factors influence people's decision-making processes? What are the benefits of, and barriers to, using privacy-preserving technologies in this context? What direction might be followed to foster adoption of such technologies? Are there cultural differences involved in the decision-making?

To start to address these questions, we conducted the qualitative study by combining the methods of focus group sessions (two groups with five participants each), individual interviews (18 participants), and a questionnaire survey (60 participants).

In this study, we want to understand how human factors can have an effect on the decision-making process pertaining to the trade-off between privacy and utility in mobile services. We are also interested in understanding whether certain privacy-preserving technologies can positively influence such a trade-off. In addition, we want to explore the impact of cultural differences.

This research presented several challenges:

1. **Privacy paradox.** Despite the various levels of privacy awareness and knowledge of participants in this study, people's stated attitudes may be influenced by the so-called privacy paradox. Therefore, participants' comments should be examined carefully with data collected from their technical background and with their observed behaviours.

2. **Cultural differences.** Privacy concerns and related approaches of protecting privacy vary between cultures. For instance, Esteve [86] suggests that protecting data privacy is a fundamental right in the EU, where the primary concern is invasion of privacy by corporations, whereas there is no such

"right to data protection" in the USA, where the primary concern is more about invasion of privacy by the government. The interview subjects of our study (28 participants in total) were drawn from diverse cultures and backgrounds (around 38% of them were British, 38% were Chinese, and 24% were participants from other countries). Hence, this factor should also be carefully taken into consideration when examining their responses.

3. **The skew of ad click-rates.** In Section 3.1.2, we introduced the core ecosystem of the trade-off between privacy and utility — free apps and highly targeted ads, which makes the click-rates of ads an important determinant of our suvery. According to the study conducted by Gill *et al.* [87], users' contribution to advertising revenue is significantly skewed — ad clicks from 20% of the users make contribution to 80% of the revenues. The result suggests that users who click on ads (actively participate in the trade-offs) only account for a relatively small percentage. Hence, the fact should be kept in mind that related bias might be introduced and the result might be influenced due to the relatively small sample size (60 in total) of our research.

4. **A complex array of factors.** People's decision-making can be affected by a complex array of factors. A great deal of research has been conducted to study such factors. For example, research conducted by Kamleitner *et al.* [88] showed that users would attach value to their personal data and many of them would make trade-off decisions by comparing the payback with their desired price. Samat *et al.* [89] studied how users' disclosure decisions could be influenced by the interplay of presentation and content. Research conducted by Bonné *et al.* [90] suggested that the factor of gender can also be involved in the decision-making process. Hence, it is important to declare that the human factors we discuss in this study cannot fully explain the decision-making process. In addition, these factors and their importance to the participants are relatively subjective.

3.2.2 Study approaches and participants

This study was conducted in Oxford, UK, in February 2017. There are two primary qualitative approaches adopted in the study: focus group interviews (10 participants) and individual interviews (18 participants). In addition, a questionnaire survey (60 participants) also provided auxiliary data.

Focus groups are useful for gathering detailed information about both personal and group feelings [91]. A broader range of information can also be collected from the interaction and discussion among the participants. By contrast, individual interviews are helpful for researchers to probe more deeply on each topic and get information from non-verbal responses [92].

Overall, the method of cluster sampling is adopted as the sampling strategy in this study. The factor of knowledge of privacy technology is the main basis in the design of the focus groups, while the cultural differences is the main consideration in the individual interviews.

To ensure proper interaction during the focus group sessions, two groups based on participants' technology backgrounds were selected for the current study. Members of the first group are young cyber security researchers, who are all doctoral students in the field of system and software security. Instead of the strong knowledge of security technology, the second group shows an international diversity: the members are adults from five different countries with different occupational backgrounds.

The participants of individual interviews were also categorised into two groups to investigate the similarities and differences between attitudes between young British citizens and young Chinese citizens. In particular, the interviews with Chinese residents were conducted via Skype.

The group size of each focus group session was 5 and the size of each individual interview category was 8. The total number of interview subjects was 26. Each of the subjects was offered an £10 Amazon gift voucher for their time. The distribution of participants is shown in Table 3.2.

Group(N)	Type	Description
Cyber security doctoral students (5)	Focus Group	Cyber security doctoral students in the University of Oxford, between 26-45
General adults (5)	Focus Group	Adults of various countries, various occupation, and various ages; two members between 26-35, three others between 36-45
Young British adults (8)	Individual Interview	Young British citizens, between 26-35
Young Chinese adults (8)	Individual Interview	Young Chinese citizens, between 26-35

Table 3.2: Participants description

3.2.3 Scenarios and data collection

Wash *et al.* [93] argued that the result reflected by users' actual behaviours is more accurate than that collected by their self-reports when studying certain user decisions. To better measure participants' actual behaviours, a card-exchanging game was applied in the focus group sessions to start off the discussion.

Participants were asked to play the roles of mobile users and service providers in a two-round card-exchanging game. Five cards representing personal information (e.g. gender, location, contacts list) were issued to "mobile users", whereas five cards representing mobile services (e.g. real-time weather forecast, recommendations of interested activities, £5 Amazon voucher as the compensation) were issued to "service providers". In the first round, "mobile users" were asked to select 0 to 5 information cards they can afford to disclose, then exchange them with the corresponding number of service cards decided by "service providers". Note that, in this round, "mobile users" were not able to figure out the service cards they could obtain before the trade-off was finally made. In the second round, "service providers" were asked to show all their service cards before performing the trade-off. Therefore, "mobile users" were able to choose the particular service cards they really want to have, and then exchange them with self-selected information cards.

The card game served as an icebreaker and helped participants to build mental associations with privacy and utility exchange. Also, participants' actual behaviours were observed during the game.

Apart from the card game, the same scenarios were applied in both the focus group sessions and the individual interview sessions to collect participants' comments.

At the beginning of each session, five scenarios that mobile users typically encountered in their everyday lives were presented. The aim was to explore participants' experience with balancing privacy and utility related to mobile services. The scenarios were as follows:

- **Scenario 1**. Experience of exposing personal information actively: imagine that you are posting stories or photos on Facebook or Instagram.
- **Scenario 2**. Experience of interacting with mobile system providers: imagine that you are downloading an app from the official Android or iOS app stores.
- **Scenario 3**. Experience of receiving targeted recommendation services: imagine that you are visiting an unfamiliar place and you receive recommendations for useful information such as tourist attractions, hotels, and restaurants on your mobile device.
- **Scenario 4**. Experience related to financial activities: imagine that you are doing online shopping on your mobile device.
- **Scenario 5**. Experience related to personal events: imagine that you are managing calendar events on your mobile device.

For each scenario, participants were requested to think about and comment on the following topics:

1. What type of personal information do you think is being collected in this scenario? (To explore participants' awareness and knowledge of privacy leakage.)
2. Who do you think is collecting your personal information? Where do you think your personal information will end up? (To explore participants' perceptions of the leakage path of their personal information.)

3. How do you think they collect or deduce your information? (To explore participants' knowledge of the different personal information-collecting technologies.)

4. In what ways do you think your personal information will be used? Do you see any potential risks? (To explore participants' awareness and knowledge of privacy risks.)

5. What utility do you think is obtained in this scenario? Do you think it is worth the cost? (To explore participants' feelings of, and attitudes toward, the trade-off between privacy and utility.)

The participants were then presented with a brief introduction to the concept of the Privacy-Preserving Targeted Mobile Advertising (PPTMA) framework [11] as an example of a privacy-preserving technology. The framework was applied to the five scenarios to showed how privacy-preserving technologies could help mobile users to take control of their sensitive information. The example also introduced the possibility of taking advantage of mobile services without compromising users' privacy.

After showing the examples, participants were requested to think about and comment on the following topics:

1. What do you see as the benefits of using such privacy-preserving technologies?

2. What do you see as the barriers to using such privacy-preserving technologies?

3. Do you think such privacy-preserving technologies could influence your decision-making process of making trade-offs between privacy and utility in mobile services?

4. Who do you think should be responsible to certify such privacy-preserving technologies do what they say?

3.2.4 Results and discussion

The focus group members, individual interview members and questionnaire survey participants show certain consistency with regards to their mobile service use. For example, about 80% of them spend more than 60 minutes on mobile services per day, with more than half of the total participants spending more than 120 minutes. The mobile phones are mostly used for services such as social networking (90%), searching for information (88%), creating and checking emails (63%) and online shopping (53%). Furthermore, most of the participants had some experience in making decisions involving the trade-off between privacy and utility in mobile services.

We now introduce our findings by examining participants' comments and by reviewing the auxiliary data collected from the questionnaire survey.

Factors contributing to the trade-off

A few of the most common human factors were observed through our study. We first introduce these factors and examine the related work for each, then discuss our findings based on each of them.

1. **Awareness and knowledge of privacy risks.** Privacy awareness and knowledge can reduce the level of trust [94] and guide users to make informed decisions about the disclosure of data [95]. Previous research has found that consumers and users have very little awareness and knowledge of privacy leakage [96–98]. Our study recruited a group of participants with relatively strong privacy knowledge to investigate the differences with regards to this factor between participants with different levels of privacy awareness.

2. **Trust in service providers.** Prior studies [67, 99, 100] have shown that users' trust of the service providers (or data collectors) can increase users' willingness of sharing their information. Instead of studying the factor of trust separately, we also study how trust can affect users' privacy awareness and subsequently influence the trade-off.

3. **Desire for mobile services.** Consumers' perceived utility of mobile services is, undoubtedly, the main reason for users to disclose their personal data. In practice, users may trade their privacy for benefits such as location-based services [101], personalized services [102], or mobile social networking services [103]. Our study investigates whether (and how) the strong desire for mobile services may overcome privacy awareness and encourage users to make the deal.

4. **Beliefs of cyber privacy.** There are many misunderstandings of cyber privacy; examples include the belief that nothing is private online [104] or the statement that "I've got nothing to hide so I don't worry about information disclosure" [105]. In this study, we investigate how such beliefs of cyber privacy can affect participants' decision-making process.

Factor 1: Awareness, knowledge, and trust

It is clear that the awareness and knowledge of privacy leakage is an important factor that strongly influences people's decision-making processes. The differences with regards to this factor between the groups was greater than expected. The findings suggest that the influences vary significantly due to the participants' various backgrounds.

Pötzsch [95] argued that privacy awareness enables people to make informed decisions about the disclosure of data. The study conducted by Van Kleek *et al.* [106] also showed that people had higher confidence in their choices with the knowledge of apps' data collection and sharing behaviours. Our findings suggests that people with limited awareness and knowledge of privacy leakage are more likely to exchange their personal information for mobile services:

> "If I know my privacy is compromised I would not make the trade-off. The key problem is that in many cases I didn't realise my personal information is being collected." (I2P6, Chinese, Female, Individual interview, Semi-skilled worker)

Questionnaire: To the best of your knowledge, which of the following information is technically possible to be collected/deduced by an app provider when you use an app? (multiple choices)

Item	Frequency (Sample = 60)
Your geographical location	54
Your phone number	44
Your contacts	38
Time of usage	37
Your email address	36
Your network operator	33
The list of apps installed on your phone	31
The operating system of your phone	31
The time and frequency of the app usage	30
Your SMSs	27
Your interests	25
Your calling history	23
Your IMEI number	19
The events of your calendar	18
The screen size of your phone	17
Your age	17
Your gender	17
The retail price of your phone	11

Table 3.3: Auxiliary data: Awareness and knowledge of privacy leakage

> "I personally feel safe. Maybe because that I don't consider myself that important or special. I can't see why they want my information. So if Google requires my location details to provide services I'll give it without thinking about the trade-off" (G2P3, Finnish, Female, Focus group, House person)

On the other hand, many interviewees stated that they refused to make such trade-offs when sensitive information is required for exchanging services.

> "Yes, I have refused to install apps that ask for very specific information. I tried to find substitutes in that case." (G2P1, British, Female, Focus group, Professional and managerial occupation)

> "I didn't pay for using the apps so I understand they need my information to make money by some means. But for most apps, I would only use them if they promise that the data would only be used in aggregate form." (G1P4, American, Male, Focus group, Cyber security doctoral student)

The findings suggests that limited awareness and knowledge of privacy leakage could positively affect the user's acceptance of mobile services with risks of disclosing personal information. Meanwhile, it is noteworthy that some interviewees acknowledged that they normally consider themselves to have serious privacy concerns in a pre-survey; however, in the subsequent card-exchanging game (personal information cards versus mobile services cards), they unconsciously ignore the privacy risks and decide to make the trade-offs easily. Such findings show that, in practice, participants' observed behaviours associated with the the privacy paradox in the context of mobile services.

In the report of a study conducted by Egelman *et al.* [97], only 22 out of 718 participants correctly identified a phishing website when their awareness was tested. A consistent result was seen in our research.

The auxiliary data of Table 3.3 shows that, while most participants could identify the risks of privacy leakage about the information with a high exposure rate from public media (e.g. geographical location, phone number and contacts), certain personal information, which seems difficult to collect (but, in fact, is not), is rarely considered by the users.

Take *"the retail price of your mobile"* at the bottom of Table 3.3 as an example: it is easy to get the information of the phone model. The phone model then leads to a precise retail price, which can be used to deduce the spending power of the phone owner. Overall, participants showed relatively low awareness of such privacy leakage in our study.

Factor 2: Trust in service providers

Earp and Baumer [107] found that, if a site or a company is well-known, consumers would be more likely to disclose information to it. Similar comments received in our study suggests that trusting the reputation of a company would weaken users' awareness of privacy risks from such a company and positively impact the adoption of the trade-off. The following representative comment indicates how *trust* can affect users' privacy awareness and subsequently influences users' decision-making process.

Questionnaire: What might encourage you to give the information to an app? (multiple choices)

Item	Frequency (Sample = 60)
In exchange for access to the main functions of the app.	45
The app provides a statement regarding how the information is going to be used.	29
If there is no substitute for the app.	23
In exchange for a useful recommendation.	17
If the data would only be used in aggregate form.	16
In exchange for valued-added service.	13
In exchange for a small discount or coupon.	6

Table 3.4: Auxiliary data: Impact of trade-off situations

> "I feel safe to use some mobile services because the companies like Google or Facebook are famous and so many people are using their services. Furthermore, the government can help to monitor them." (I2P4, Chinese, Female, Individual interview, House person)

However, the extent that *trust* affects *privacy awareness* may vary according to users' knowledge of privacy risks. Although a good reputation is viewed positively by participants with more privacy knowledge, such participants show stronger concerns under the same situation.

> "I think I'm more scared of potential recruiters or potential employers looking at my profile on Facebook and use it to take advantage of me. So I never post anything public. [...] We also know no matter I post it in public or private Facebook is going to get the information. I believe in Facebook as a company, but what about their employees. We all know stories like software engineer may hack into company's servers to steal others' personal information, which happened in Google previously. In addition, company like Facebook may share our profile with countless cooperative advertisers. Can we trust these advertisers as well? All these factors make me feel hesitate to use their services." (G1P5, Chinese, Male, Focus group, Cyber security doctoral student)

Factor 3: Desire for mobile services

As we discussed in the previous subsection, *awareness and knowledge of privacy risks* clearly influences people's decision-making process. We predicted that, in general, a user with a relatively weak awareness of privacy risks is more likely to 'make the deal'. However, our findings suggest that a strong desire for specific mobile services may also enable users with strong knowledge of privacy leakage

to bear the risks. For example, a statement made by a participant shows the ineluctability of using mobile service such as Facebook:

> "I rarely post pictures or comments to Facebook and I'm careful with what I post and where from. However, as a small business, (to advertise my business) I have no choice about being on Facebook. If I didn't have to be I would avoid it." (I1P8, British, Female, Individual interview, Professional and managerial occupation)

Another statement made by a cyber security doctoral student with relatively strong knowledge of privacy risks also indicates how the *desire for mobile services* may work in this context.

> "It's such pain to type in my address or my credit card details every time to shopping online. So I use Amazon app all the time. Because it already got my credit card details, already got my post address. If my friend recommends me a book I can buy it with one click and get it by tomorrow. I know Amazon is collecting my information, that's fine. I chose to have a business relationship with them. I'm happy that they have my address, my credit card details, my purchasing habits. Because I got good recommendation and also get cheaper items from them. I feel like they have more duty not to sell my information because they already got money from me." (G1P2, British, Male, Focus group, Cyber security doctoral student)

These statements suggest that, in some situations, the strong desire for mobile services may overcome the awareness of privacy risks and encourage users to disclose their information. For example, there is no substitute for the Facebook app, and the Amazon app may provide significant benefits — such as useful recommendations. The auxiliary data of Table 3.4 shows some factors that might encourage users to give personal information to an app.

Factor 4: Beliefs of cyber privacy

In addition to the particular situations introduced in Table 3.4, there is another encouraging factor that can be easily noticed: the common belief that cyber privacy does not exist or that "I've got nothing to hide".

Phelan *et al.* [104] found that the belief that nothing is private online is often used by people as a reference frame to assess marginal risk in the context of privacy

concerns related to online data collection. A signficant number of participants made similar statements in our study.

> "The services is worth the cost. Because, in fact, there is no so-called 'privacy' in the cyber world nowadays. We're going to give out our privacy anyway, why not exchange it for some benefits?" (I2P1, Chinese, Male, Individual interview, Professional and managerial occupation)

> "Our personal information is not only exposed by mobile. From this aspect, we don't need to feel so sensitive. We are surrounding by internet, who can escape?" (I2P7, Chinese, Female, Individual interview, Semi-skilled worker)

> "There are many ways to use our information. Maybe to improve the service to make you use it more or maybe they just want to sell it or use with advertisers. If we think too much we'll have no app to use, because every app is doing this." (I1P6, British, Male, Individual interview, Student)

The responses of participants indicate that they could achieve low privacy concern even if subjectively they would like to pay more attention to their personal data.

Desired adoption of, and trust in, the privacy-preserving technologies

Different types of privacy-preserving technologies have been proposed to help users handle the privacy and utility dilemma [11, 41, 81, 108]. While the theoretical advantages of such tools are clear, little adoption has been observed in practice [109, 110]. Some researchers have already started to explore the direction that might be followed to foster adoption of privacy-preserving technologies [111]. Some others [112, 113] provide foundations for future research on designing incentive schemes for privacy-sensitive users by proposing models to capture the interactions between businesses and consumers in the context of targeted advertising. In this study, we explore participants' attitudes of importing privacy-preserving technologies into this context.

Not surprisingly, participants' responses suggest that they were in favour of privacy-preserving technologies that can provide utility for users to manage their personal information in mobile devices, and can help users to increase their awareness when using mobile services. Such technologies have the potential to

decrease users' privacy concern and establish a win-win situation for both users and mobile service providers.

> "Yes, if I thought any technology could give me better service as a result of more fine-grained information I'd probably agree that such technology would influence my decision making process." (G2P4, Korean, Female, Focus group, white collar worker)

> "Such technologies would make me more likely to download an app if coarse-grained data was collected only." (I1P4, British, Female, Individual interview, Student)

In addition, the same participants' different behaviours observed during the two round card-exchanging game also echo the adoption of personal information management tools. Compared to the first round of the game, participants suggested that in the second round, when they could clearly figure out what services they might obtain and when they were able to decide which part of their personal information should be submitted to the service providers, a better decision could be made.

While the necessity for privacy-preserving technologies is clear, participants expressed their concerns about the potential trouble caused by adopting such technologies. For example, some participants worried about the costs of learning how to use such tools. Some provided comments pertaining to their previous experience of receiving disturbing notifications caused by uninterrupted reminders from such tools.

> "It's useful. But I think it's difficult for my parents to learn how to use it. They are not good at IT stuff." (I2P6, Chinese, Female, Individual interview, Semi-skilled worker)

> "I get a bit annoyed once you use some apps and then you start to get notifications (from the privacy-preserving tools) reminding you to protect your data." (G2P5, Australian, Male, Focus group, white collar worker)

Participants also provided comments on basic requirements for decreasing such barriers, including configurability, easy deployment, privacy by default, and automatic activation.

Apart from these requirements, there is a major concern about trust in such tools. Many people are still confused about the monitor in this context: who is responsible

to certify such technologies do what they say? What if a privacy-preserving technology provider steals users' information with the excuse of protecting it?

> "It's tricky with electronic stuff. What's that to do with my data? [...] If you have that stamp, then we know you follow the rules. That might be easier. But whose job is it (to give the stamp)?" (G2P2, Chinese, Female, Focus group, House person)

The overall results of the responses are shown in Table 3.5. A noteworthy percentage (43%) of the participants cannot figure out a suitable entity to play this role. Most of the rest rely on a government department to be the supervisor:

> "Government should publish related policies, and companies with high reputation should help to implement the policies and monitor related technologies." (I2P5, Chinese, Female, Individual interview, Semi-skilled worker)

In the meantime, some participants also show different preferences on this:

> "Maybe the government but can they be trusted? I hope that third party standards exist but I'm not sure if they do and how or who could regulate them." (G2P1, British, Female, Focus group, Professional and managerial occupation)

> "To some extent the person downloading the app is responsible for checking what data will be taken from them, because when he accesses the (Android) app store, the app to be downloaded has already told the him what permission it needs and what information it will collect." (I1P4, British, Female, Individual interview, Student)

In addition, responses on this topic also showed the characteristics of different cultures, which will be discussed in the next section.

Privacy concerns across cultures

Kayes *et al.* [114] argue that users' privacy concerns vary between cultures. Specifically, users from individualistic countries give more weight to privacy protection compared to those from collectivistic countries. Our study also explored the influences of cultural differences regarding the protection of privacy.

The first interesting result emerged when we review the responses associated with "Trust in the privacy-preserving technologies". While more than half of Chinese

Questionnaire: Who do you think is responsible to certify such technologies do what they say? (multiple choices)

Item	Frequency (Sample = 60)
Government department	26
Don't know	26
Organisations	8
Famous companies	4
Others	1

Table 3.5: Auxiliary data: Trust in the privacy-preserving technologies

participants (22 out of 43) stated that a government department should play the role to monitor technologies providers, the corresponding figure for participants from western countries was about 24% (4 out of 17). From a different aspect, our findings suggest that users from collectivistic countries are more likely to expect a government department to protect their individual privacy, while users from individualistic countries are more likely to emphasize the roles of individuals and non-governmental organisations. Representative comments from each groups are as follows:

> "I go for the government. Because companies and organisations may take advantage of the monitor position for their own benefits. [...] (when asked about the possibility of government surveillance) Maybe, but it's better to put my information in the hands of the government rather than in those of some companies." (I2P2, Chinese, Female, Individual interview, White collar worker)

> "I rely on information from groups such as ORG (Open Rights Group, a UK-based organisation that works to protect the rights to privacy and free speech online) at the moment, but all options such as government, organisations or famous companies are problematic." (I1P8, British, Female, Individual interview, Professional and managerial occupation)

We also saw some similarities between participants from different cultures. For example, with regards to the open-ended question in the questionnaire survey "Who is collecting your information? Where will the information end up?", most of the participants from western countries and China both mentioned advertisers, software developers, social network platform such as Facebook and Instagram (mentioned by participants from western countries), or Weibo (mentioned by Chinese participants). However, surprisingly, no one mentioned the government among all 43 Chinese

Factors	Contributing items	Effect on decision-making
Awareness and knowledge of privacy risks	high with unreliable services require sensitive information	-
	high with reliable services	+
	low	+
	paradox	+
Trusting on the reputation of a service provider	high	+
	low	-
The degree of desire for mobile services	high	-
	low	+
The belief of cyber privacy	"Nothing is private online" or "I've got nothing to hide"	+
	"Privacy can be preserved online"	-
Impact situations (examples)	App provides decent main functions	
	Apps clearly state how the information is going to be used	+
	No substitute for the apps	+
	Apps provide useful recommendation	+
	The data would only be used in aggregate form	+
	Apps provide valued-added service	+
	Apps provide small discount or coupon	

Table 3.6: factors influencing decision-making pertaining to the trade-off

participants in this question, whereas 2 of the 17 participants from western countries proposed the possibility of government surveillance.

It is important to declare that the sample size of participants was relatively small and the comments made by participants were relatively subjective. However, to a certain extent, our finding suggests that the Chinese participants tend to rely on the government to protect their privacy and mainly care about the invasion of privacy by big corporations. The participants from western countries, by contrast, were concerned about the disclosure of personal information to corporations, as well as the government.

3.2.5 Lessons learned from the user study

Evidence from the participants' comments suggests that users were in favour of the various benefits provided by mobile computing systems. It is clear that, in order to take advantage of such systems appropriately, many human factors should be take into consideration.

Our goals in this user study were: (1) to investigate the human factors influencing decision-making pertaining to the trade-off between privacy and utility in mobile services; (2) to explore the direction that might be followed to foster adoption of privacy-preserving technologies that could help mobile users to balance the privacy and utility in more effective ways; and (3) to study the influences of cultural differences regarding the protection of privacy.

The findings suggested that human factors such as users' awareness and knowledge of privacy risks, users' trust of service providers, users' desire for mobile services, and their belief of cyber privacy are all important contributing items of the decision-making process. Meanwhile, users are also encouraged to share their information in particular situations. These findings are summarised in Table 3.6.

In addition, most participants expressed a strong desire for adopting privacy-preserving technologies in this context. While there is a potential to establish a win-win situation for both users and mobile service providers, participants also showed their concerns about the potential trouble that might be caused by adopting such technologies, including learning cost, frequent reminders, and a major concern — the need for a governing body to monitor the technology providers.

With regards to the governing body, most Chinese participants stated that the government department should play the role, whereas this approach was less popular with the UK participants.

Theoretical implications

It should be emphasised that this study can claim limited novelty with regards to implications; instead, this study confirms several theoretical contributions which echo findings from the existing research on privacy issues of mobile services.

First, this research investigated the human factors that influence the decision-making process pertaining to the trade-off between privacy and utility in mobile services. This advances our understanding of mobile users' behaviour and the potential causes of the privacy paradox. Future research can expand and weigh the factors to build related models.

Second, our findings suggested that, although users are aware of their privacy leakage, an appropriate compensation mechanism could still encourage them to get involved in the privacy-trading process. To this end, if users are given corresponding rewards, as well as the ability to make independent choices with regards to releasing specific parts of their personal information, they are more likely to adopt the exchanging relationship with service providers. From this perspective, future research can explore potential strategies to provide mutual benefits for both sides of the trade.

Third, this research enhanced the theoretical understanding of the differences of privacy concerns from cross-cultural dimensions. Future research can be conducted to further study the effects of culture on privacy concerns.

Practical implications

This study provides guidelines for mobile service providers that consider improving privacy trust to recruit more privacy-sensitive users. It also shows the direction that might be followed to foster adoption of privacy-preserving technologies to help mobile users balance privacy and utility.

First, practical suggestions for mobile service providers or app developers can be provided by examining the factors discussed in this study. For example, strategies such as providing a statement regarding how users' information is going to be used in apps, using data in only aggregate form, or providing valued-added service and coupons can effectively encourage users to release the information and apply the related mobile services.

Second, the findings suggested that more attention should be paid to enhancing users' trust in privacy-preserving technology. Although the benefits of applying such technologies are clear, users' distrust can negatively affect the adoption.

Third, our findings suggested that, pertaining to the role of supervising privacy-preserving technologies, none of the government departments, non-governmental organisations, or large companies can alone obtain users' trust. A hierarchical mechanism that involves all of the stakeholders is worth considering in practice.

Limitations

It should be emphasised that, as with any empirical research, our study has limitations and certain bias. The results are limited due to the sampling methodology adopted and the relatively subjective factors involved. First, the sample size of our research is relatively small and the proportion is not well-balanced. The total sample of this research is 60, with 17 participants from western countries and 43 participants from China. Second, the main participants in our research are young adults — most of whom had received undergraduate or post-graduate education. As users with lower education level and users in other age ranges are also important mobile users, an extended range of samples should be applied in the future study. Third, the factors and their importance provided by participants are relatively subjective. A larger sample in future work can help to mitigate the subjective influences. Additionally, as discussed by Acquisti *et al.* [19], factors such as incomplete information, information asymmetries, users' inability to acting optimally on vast amounts of data etc., are all able to influence the decision-making process, which could, in turn, import bias and limitations into our study.

3.3 Requirements and threats

Motivated by the results of the user study, we outline a list of high level requirements for the enhanced TMA system. The goal of the improvements is to form a privacy-preserving system that has the potential to help users handle the privacy and utility dilemma of mobile services, with particular consideration given to TMA services.

3.3.1 Scope

As mentioned in Chapter 1, our research question is:

> *Is it possible to develop a privacy-preserving TMA framework that enables mobile users to take advantage of useful advertising services without their privacy being compromised, and without impacting significantly advertising effectiveness?*

In particular, by "*useful* advertising services", we mean those services with relatively high acceptance rate, including: free, helpful information (e.g. location-based recommendation), money-saving information (e.g. coupons or discounts), personalised entertaining information (e.g. recommendation of personalised news or products), and information that can enhance consumers' sensual and aesthetic experiences of brands (activities of luxury brands).

By "*their privacy*", we refer to the personal data stored in mobile devices and related users' preferences that can be deduced based on such data. The personal data involved includes dynamic information such as location information and static information such as unique device ID. A detailed description of such personal data was described in Section 2.2.2.

By "without impacting *significantly* advertising effectiveness", we refer to those changes that require only slight or no modification of ad-networks' basic framework or infrastructure; that only slightly affect the performance of delivering useful advertising services with unnoticed changes; and that could keep, if not improve, the click-through rate of the targeted ads.

With respect to particular cultures, we will identify and discuss the differences when applying the improvements in different cultures. However, the basic idea of our solution should be universally beneficial.

The design and implementation of our solution will be based on the Android system. However, the core mechanisms of the solution should be transferable to iOS and other mobile platforms.

3.3.2 High level requirements

Inspiring by the feedback collected from the user study, we outline the following high level requirements for the enhanced TMA system:

1. **Privacy-preserving**. The improvements should allow mobile users to benefit from useful targeted mobile ads without there being a concern that their personal information will be leaked. The ideal solution should store and

manage all kinds of users' privacy data inside the mobile device, rather than disclosing the information.

2. **Ad scanning and verifying**. The improvements should be able to help users identify ads or malicious behaviours. In particular, the improvements should be able to scan apps and discover the advertisement plugins that they contain and monitor the behaviour of apps to determine whether they are behaving legitimately.

3. **Controlled access**. The improvements should enable the control of access to users' personal information. Ideally, they should provide fine-grained control over what kinds of data are made available to third-parties, rather than applying the default permissions system provided by mobile platforms.

4. **Data management**. The improvements should allow users to manage their personal information manually, meaning that they can add, delete or update the particular personal information, as well as save different copies.

5. **Business supporting**. A TMA system involves many different stakeholders. To maintain an effective ecosystem, the improvements should also take into account the benefits of the other participants apart from users. Ideally, the improvements should serve as an interface for commercial organisations that would be happy to provide targeted ads without holding and controlling users' personal information.

6. **Appropriate governance**. While the improvements could enable mobile users to balance privacy and utility in a more effective way, there is the need for a governing body to detect privacy violations made by the privacy technology providers themselves. Hence, the improvements should allow privacy-advocates or governments to monitor the privacy improvements appropriately. For instance, functions related to data-sharing might be able to be separated and run on servers governed by different parties, such as privacy technology providers, privacy-advocates and governments.

3.3.3 Threat models

Four main kinds of participating stakeholders of the TMA system are identified in Chapter 1: users; ad-networks; advertisers; and content providers. We now present threat models from the different perspectives of each.

Perspective of users

Other than the privacy-friendly ad-networks, who would like to provide privacy-aware advertising services, most of the other stakeholders — advertisers, content providers, and general ad-networks — can take potential adversarial roles with respect to the users. We assume that these adversaries would try to collect users' personal information — secretly or compulsorily — in order to perform TMA. Thus, we consider users' main adversaries to be the ad-networks that are plugged inside apps and collect users' personal information based on users' interactions with the involved apps.

Another class of adversary is untrusted third-parties who try to collect users' information and trade it with others. These untrusted parties also perform the collection through apps that require related permissions. Therefore, the system should monitor the behaviour of malicious apps even though they do not contain any ad-network plugins.

Perspective of ad-networks

The improvements would allow mobile users to identify apps with ad-plugins and revoke sensitive permissions from these apps. The fact that mobile users could make use of these features to block ads from particular ad-networks or apps would make ad-networks unhappy. Thus, from the perspective of ad-networks, such users could be their adversary.

However, ad-networks should acknowledge that, even in the setting for Google Ads, users can choose to opt out of seeing targeted ads. Apple also provides an ad-blocking feature in iOS 9 for users to block ads on their mobile devices. Therefore, a better way to establish a healthy environment is to give users the choice to opt

out of the targeted advertising services if they would like to. Treating such users as adversaries could, in turn, drive ad-networks to improve the efficiency of ad-delivery and to explore privacy-aware approaches for performing targeted advertising.

Another common issue concerned by ad-networks is click-fraud. Click-fraud could be performed by advertisers to attack competitors, or performed by content providers to earn extra income. Thus, we consider people who perform click-fraud as another class of adversary. Therefore, a mechanism for click-fraud detecting is required.

Perspective of advertisers and content providers

Ad-networks charge money from advertisers for delivering their ads to users and share the payment with content providers for displaying ads on their apps or websites. The amount of the charged money and that of the shared payment is calculated based on the view/click reports. Hence, both advertisers and content providers may consider untrusted ad-networks as adversaries — who might present (fake) higher click reports to advertisers to charge more money and present (fake) lower click reports to content providers to share less. Apart from the potential fake reports provided by ad-networks, advertisers are also under threat from click-fraud attacks from other entities such as their competitors or the content providers who display their ads. Therefore, a mechanism to provide authentic view/click reports to all kind of stakeholders is required.

Additionally, users who block ads may also influence the interests of advertisers and content providers. As discussed above, this could, in turn, drive advertisers to improve the quality of their ads and drive content providers to explore the potential for improving the user experience with regards to displaying ads.

3.4 Summary

In this chapter we have discussed people's perspectives on privacy issues pertaining to mobile services. The results of a user study on the trade-off between privacy and utility have been reported. Based on the results, we have discussed the goal, the high level requirements, and the threat models of a privacy-preserving system that has

Stakeholder	Threats
Users	1. Ad-networks collect users' personal information with ad-plugins. 2. Untrusted third-parties collect (and trade) users' information with apps holding related permissions.
Ad-networks	1. Users block ads from particular ad-networks or apps. 2. Competitors or content providers perform click-fraud attacks.
Advertisers	1. Ad-networkers provide fake view/click reports. 2. Competitors or content providers perform click-fraud attacks. 3. Users block ads from particular ad-networks that the advertisers cooperate with.
Content providers	1. Ad-networkers provide fake view/click reports. 2. Users block ads from particular ad-networks that the content provider cooperate with.

Table 3.7: Threats

the potential to help users in their decision-making. In the following chapter we will show how formal models might be used in both helping to reason about the balance between privacy and utility and in helping to underpin the design of our solution.

4

Formal models: From TMA to PPTMA

Contents

4.1 Formal models in the context of TMA 70
4.1.1 Formal models and privacy 70
4.1.2 Purposes of the formal models 71
4.2 From TMA to PPTMA 73
4.2.1 An abstract TMA system 73
4.2.2 PPTMA: A privacy-preserving solution 74
4.3 A model of TMA 75
4.3.1 Overview of the model 76
4.3.2 Privacy-affected operations 81
4.4 A model of PPTMA 85
4.4.1 Overview of the model 85
4.4.2 Improved operations 89
4.5 Application of the PPTMA model 92
4.5.1 The first stage of the ad-selection process: pre-download ads 93
4.5.2 The second stage of the ad-selection process: local ad selection 95
4.5.3 Click-audit obfuscating and click-fraud detecting 96
4.6 Analysis 98
4.7 Summary 100

Having discussed the goal, the high level requirements, and the threat models of a privacy-preserving TMA system that has the potential to help in striking a balance between privacy and utility in the previous chapter, we now give consideration to

how formal models can help in underpinning such a system, in analysing privacy in the context of TMA, and in allowing users to specify control of their personal information. We start by discussing the roles that formal models can play in reasoning about privacy in general and specifically in the context of TMA.

The contents of this chapter are based on *Yang Liu and A. C. Simpson: Privacy-preserving targeted mobile advertising: Formal models and analysis. In: Proceedings of the 11th International Workshop on Data Privacy Management (DPM 2016). Lecture Notes in Computer Science Volume 9963, Pages 94 — 110, 2016* [112].

4.1 Formal models in the context of TMA

4.1.1 Formal models and privacy

As mentioned in previous chapters, TMA is an important part of the internet economy with inherent conflicts of interest among the stakeholders. Our focus is a solution that tries to steer a middle path for users and corporations and that has the potential to be palatable to both of them. To this end, we have proposed high level requirements for a *Privacy-Preserving Targeted Mobile Advertising* (PPTMA) system with a view to users taking advantage of targeted ads without their privacy being compromised and organisations benefiting from higher response rates than would be possible via a solution that took a more anti-corporate stance.

As argued by Tschantz and Wing [115], formal models have many roles to play in reasoning about privacy in a variety of contexts. As an example, in [116], Abe and Simpson illustrate how formal models can be helpful in providing assurances of privacy in the context of data sharing. In terms of "privacy-specific needs", Tschantz and Wing argue the following:

> "We want to allow users to control how much of their information is released to others, but we want to make it easy for them to specify this control, and even more challenging, to understand the implications of what they specify and to be able to change the specifications over time." [115]

The contribution described in this chapter is in the spirit of that argument.

Formal models can be beneficial in many ways. In this chapter, we give consideration to how formal models can help in underpinning a proposed system of *Privacy-Preserving Targeted Mobile Advertising (PPTMA)*, in analysing privacy in the context of TMA, and in allowing users to specify control of their personal information. The underpinning models have been developed in terms of the schema language of Z [117]. The Z notation has been used due to its accessibility: it is widely taught and its structures have much in common with those of the relational model of data. In addition, Z has good tool support in the form of ProZ [118], which supports both animation and model-checking.

4.1.2 Purposes of the formal models

Having discussed the beneficial roles that formal models can play in the context of TMA, we now introduce the main purposes that formal models can serve with regards to our solution.

Recalling our research question introduced in Chapter 1, our main aim is to provide a solution that is able to:

1. *Maintain, if not enhance, the mobile advertising effectiveness for corporations, and reduce users' hostility.*

2. *Preserve privacy for mobile users, and enable them to take advantage of useful advertising services.*

A number of contributions (such as [24], [14], [35] and [3]) have been proposed in the same spirit. However, as discussed in previous chapters, while the theoretical advantages of privacy-preserving tools are clear, little adoption has been observed in practice.

Our user study introduced in Chapter 3 indicates that the privacy knowledge and awareness can make positive influences on people's decision-making. In the context of TMA, the knowledge of apps' data collection and sharing behaviours can also provide confidence to users and help to improve the adoption of the privacy-preserving technologies.

This gives rise to a further aim:

3. *Provide a solution that makes the control of personal information easily specified by mobile users, and enables users to understand the effects of their decisions.*

These aims represent the primary motivation for the model. In turn, the models give confidence in our proposed solution and help to underpin the decision-making process (both in terms of ad selection and in terms of access control). To this end, the formal models serve the following purposes.

1. They help to reason about the balance between potential benefits.
2. They help to provide assurance with respect to the preservation of privacy when using TMA, and to measure the balance between utility and privacy.
3. They help support the access control decision-making process, allowing users to understand (and, to an extent, specify) how much of their personal information is shared.
4. They underpin the ad-selection process, so that it takes into account a wide range of data — but only data that users have granted access to.

In the remainder of this chapter, we will present our models in stages. In Section 4.2, we first abstract the typical workflow of a TMA system and then introduce how the PPTMA system can help in addressing the privacy issues involved in the workflow. Then, in Section 4.3, we start by formalising important aspects of current TMA systems which enables us to identify the features of such systems that can impact users' privacy. Then a further specification that describes our solution and helps to underpin our design is introduced in Section 4.4. Then, in Section 4.5, the final model is applied in a mainstream ad-selection mechanism to show how the balance between utility and privacy can be reasoned about, and how users can understand (and, to an extent, specify) the extent to which their personal information is shared.

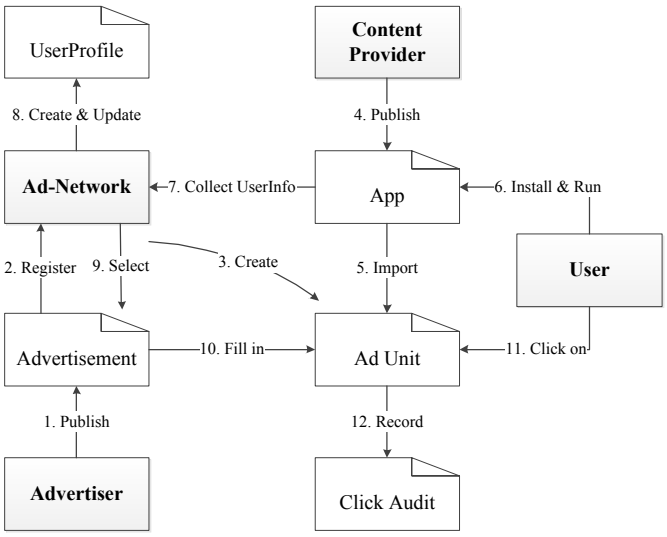

Figure 4.1: A typical TMA workflow

4.2 From TMA to PPTMA

We have briefly introduced the core concept of the TMA ecosystem in Chapter 2. In this section, we first present a more detailed workflow of such systems with the privacy-related steps being indicated. Then we introduce the key features of the PPTMA system and explain how these features can address the privacy issues involved in the TMA workflow.

4.2.1 An abstract TMA system

A TMA system automatically selects ads that are most relevant to the target user's profile and then presents those ads on their mobile device. There are four main kinds of actors in the system: advertisers, who publish ads for their products; content providers, who place those ads in their own apps; ad-networks, who collect ads from advertisers and serve them to content providers; and users, who interact with their mobile devices and click on ads.

Figure 4.1 shows key elements of an abstraction of current TMA systems, together with a typical workflow (consisting of five phases) between these actors:

1. An advertiser publishes a new ad and registers it to an ad-network; meanwhile, the ad-network creates some ad units for registered ads (steps 1–3).

2. A content provider develops a new app, and imports ad units into the app for the ads to be displayed (steps 4 and 5).

3. A user installs the app onto a mobile device and runs it, the app then collects the user's personal information (by calling mobile sensors, checking stored data, recording the user's operations, etc.) and submits it to the ad-network (steps 6 and 7). The ad-network then creates a user profile to track the user's interests, and regularly updates it with new personal data (step 8).

4. With user profiles, the ad-network selects the most relevant ads for particular users and fills ads into ad units in the active app (step 9 and 10).

5. If the user is interested in the displayed ads and performs click operations, the operations are recorded as click-audits by the ad-network. The audits can then be used as references for charging money from the advertiser and for sharing the payment with the content provider (steps 11 and 12).

Users' personal information is mainly collected and analysed in Phase 3 (steps 6–8) and Phase 5 (steps 11 and 12), while the process of targeting is handled in Phase 4 (steps 9 and 10). Our prototype solution and related models focus primarily on the privacy issues involved in these phases.

4.2.2 PPTMA: A privacy-preserving solution

To address the privacy issues exposed in the TMA workflow, a privacy-preserving solution is proposed to refine the TMA system into a *PPTMA (Privacy-Preserving Targeted Mobile Advertising)* system. We now briefly consider the basic features of the PPTMA system.

At a high level, PPTMA is a service-based solution that works as a piece of middleware positioned between untrusted third-party apps and the underlying database on mobile systems. The service runs in the background of the system, and serves the following key functions.

1. **Personal data management.** Users' personal data can be managed manually with PPTMA. The system enables users to create different copies of their particular personal information and edit them separately.

2. **Access control.** A fine-grained access control mechanism allows users to decide what kinds of data or which copies of their personal information can be made available to third-parties. Users can customise the data that is collected in Phase 3 (steps 6–8) of the TMA workflow of Figure 4.1.

3. **Ad-network scanning and verifying.** Users can scan all apps installed on their mobile devices to discover the ad plugins as well as the related ad-networks involved in the apps. Thus, users can identify specificed apps that need particular data control strategies. It also provides a service to monitor the behaviour of installed apps to detect malicious third-parties.

4. **Local ad selection and click-audit obfuscating.** PPTMA can serve as a local TMA system that performs ad selection on mobile devices: personal information is stored and analysed on mobile devices, rather than submitted to the servers of ad-networks. In addition, click-audit information that helps to trace users can be obfuscated in PPTMA before being submitted to ad-networks. The features addresses the privacy issues involved in Phases 4 (steps 9 and 10) and 5 (steps 11 and 12) of the workflow of Figure 4.1.

Functions 1 and 2 enable users to take control over their personal information; functions 3 and 4 offers a way of serving targeted ads without users' personal information being collected.

4.3 A model of TMA

Based on the workflow we discussed above, we now present a brief overview of the formal model of typical TMA systems to identify privacy-related behaviours.

4.3.1 Overview of the model

A typical TMA system is built up from many smaller components. In order to make our specification easier to comprehend, we identify and describe the components separately in five subsystems, and then combine them. To this end, we present the possible states of the following subsystems respectively. For the sake of brevity, we have omitted type definitions.

1. *ActorSystem* maintains information pertaining to the four kinds of actors: advertisers, ad-networks, content providers, and users.

   ActorSystem
   $advertiser : AdvertiserId \pfun Advertiser$
   $adNetwork : AdNetworkId \pfun AdNetwork$
   $contentProvider : ContentProviderId \pfun ContentProvider$
   $user : UserId \pfun User$

   Here, identifiers are mapped to elements of respective types.

2. *AdSystem* is concerned with publishing and registering new ads. Newly published ads should be set to a particular format, assigned to target audiences, and associated with one or more keywords and categories.

   UserBasicInfo
   $gender : Gender$
   $age : Age$
   $location : Location$
   $language : Language$

   AdFormat
   $type : Type$
   $style : Style$
   $size : Size$

   Ad
   $format : AdFormatId$
   $targetAudience : \mathbb{P}\ UserBasicInfo$
   $keyword : \mathbb{P}\ Keyword$
   $category : \mathbb{P}\ AdCategoryId$

AdUnit

$$format : AdFormatId$$
$$adNetwork : AdNetworkId$$

AdSystem

$$ad : AdId \pfun Ad$$
$$adNetwork : AdNetworkId \pfun AdNetwork$$
$$adCategory : AdCategoryId \pfun AdCategory$$
$$adUnit : AdUnitId \pfun AdUnit$$
$$adFormat : AdFormatId \pfun AdFormat$$
$$adUnitInAdNetwork : AdUnitId \pfun AdNetworkId$$
$$adInAdNetwork : AdNetworkId \pfun \mathbb{P}\, AdId$$

$$\mathrm{dom}\, adInAdNetwork \subseteq \mathrm{dom}\, adNetwork$$
$$\forall ids : \mathbb{P}\, AdId \bullet ids \in \mathrm{ran}\, adInAdNetwork \Rightarrow ids \in \mathrm{dom}\, ad$$
$$\mathrm{dom}\, adUnitInAdNetwork \subseteq \mathrm{dom}\, adUnit$$
$$\mathrm{ran}\, adUnitInAdNetwork \subseteq \mathrm{dom}\, adNetwork$$

Each ad contains the following attributes, which are described in the above schema.

(a) Format. In our model, each format of ad consists of three parts: type, style, and size. Basic types of ads could be raw text or static image. Additionally, ads can be referred to other complicated types such as rich media, flash, video, audio, animated image, link units, and so on. Ads in different styles can be referred to banner ads, app-wall ads, interstitial ads, etc. In addition, each ad is assigned to a particular size. For instance, the size of a banner ad could be '320x50 density-independent pixels' as a normal banner or '468x60 density-independent pixels' as a larger one. In addition, it might also be set to some dynamic sizes — smart-banner-size for example, which dynamically adjusted the ad to full-width and auto-height to adapt different devices. According to the formats of the registered ads, relevant ad units are created and assigned to apps with ad-plugins.

(b) Target audience. Each ad can be assigned to one or more targeted user groups (described in *UserBasicInfo*) with particular gender, age, location, language, etc.

(c) Category. Each ad can be related to one or more categories. For instance, an ad of basketball shoes can be referred to the categories of *Sport*, *Shoes*, and *Men's* separately or simultaneously.

(d) Keyword. Similar to categories, one or more keywords can be assigned to an ad at the same time.

The subsystem *AdSystem* then describes how the ads, categories, ad units, and formats are managed. In particular, the last function *adInAdNetwork* describes which ads have been registered to which ad-networks. Additionally, it tells us that one ad can be registered to more than one ad-networks at the same time.

3. *AppSystem* models the system for content providers to publish new apps and register their apps to particular ad-networks by importing related ad-plugins.

AppSystem

$$
\begin{array}{l}
app : AppId \pfun App \\
adUnitOfApp : AdUnitId \pfun AppId \\
appInterest : AppId \pfun \mathbb{P}\, AdCategoryId \\
\hline
\mathrm{ran}\, adUnitOfApp \subseteq \mathrm{dom}\, app \\
\mathrm{dom}\, appInterest \subseteq \mathrm{dom}\, app
\end{array}
$$

App, as the main medium for displaying ads on mobile devices, receives ads from ad-network by imported related ad units. Ad units, as described above, are units that are dispatched to apps by ad-networks, which are filled with selected ads in particular types, styles, and sizes.

The function *adUnitOfApp* describes which ad units have been imported to which apps. The function *appInterest* indicates the association between an app and one or more ad categories. For instance, an app of basketball game

can be associated with the categories *Team sports*, *Basketball*, and *Video game* in Google Ad Interest Categories [119].

4. *ProfileSystem* models how ad-networks collect users' personal data, create profiles for them, deduce their interests, etc.

$$
\begin{array}{|l}
\textit{UserProfile} \\
\hline
\textit{userBasicInfo} : \textit{UserBasicInfo} \\
\textit{searchBrowseInfo} : \mathbb{P}\, \textit{SearchBrowseInfo} \\
\textit{selfMadeDocument} : \mathbb{P}\, \textit{SelfMadeDocument} \\
\hline
\end{array}
$$

$$
\begin{array}{|l}
\textit{ProfileSystem} \\
\hline
\textit{userProfile} : \textit{UserProfileId} \pfun \textit{UserProfile} \\
\textit{userInterest} : \textit{UserProfileId} \pfun \mathbb{P}\, \textit{AdCategoryId} \\
\textit{profileOfUser} : \textit{UserProfileId} \pfun \textit{UserId} \\
\textit{profileInAdNetwork} : \textit{UserProfileId} \pfun \textit{AdNetworkId} \\
\hline
\mathrm{dom}\, \textit{userInterest} \subseteq \textit{userProfile} \\
\mathrm{dom}\, \textit{profileOfUser} \subseteq \textit{userProfile} \\
\mathrm{dom}\, \textit{profileInAdNetwork} \subseteq \textit{userProfile} \\
\hline
\end{array}
$$

UserProfile stores the personal data of related users. The involved data includes basic user information such as gender, age, location, language (described in *UserBasicInfo*), and behaviour information such as search and browsing history (described in *SearchBrowserInfo*), or self-made documents such as email content or calendar events (described in *SelfMadeDocument*).

ProfileSystem describes the system for ad-networks to maintain user profiles. The function *userInterest* indicates the association between a copy of a user profile and one or more ad categories. The function *profileOfUser* indicates which user's personal data is stored in which profile, and *profileInAdNetwork* tells us which particular user profile is created by which ad-network.

One user can be referred to more than one user profiles in different ad-networks, and each ad-network could maintain more than one user profiles for different users. Particularly, the last constraint indicates that one user can refer to no more than one user profile in each ad-network.

5. *ClickAuditSystem* records all users' click operations (including view operations for some ad-networks) on ads. Ad-networks make use of the records to settle accounts, and to update relevant users' behavioural profiles.

ClickAudit

$userId : UserId$
$adId : AdId$
$adUnitId : AdUnitId$
$date : Date$

ClickAuditSystem

$clickAudit : ClickAuditId \pfun ClickAudit$
$clickAuditInAdNetwork : ClickAuditId \pfun AdNetworkId$

$\mathrm{dom}\ clickAuditInAdNetwork \subseteq \mathrm{dom}\ clickAudit$

Each record of click audit can be used to verify the following information: the user who clicked on an ad (*userId*), the particular ad that was clicked (*adId*), the involved app or content provider (can be deduced with *adUnitId*), and the date of the click operation.

In addition to the function of verifying, the click record can also be used to update a user profile — as the click on a particular ad might indicate the user's certain interests on the categories of the clicked ad.

Combining these subsystems, we define a TMA system thus (for the sake of readability, the constraints in this schema are omitted).

System

ActorSystem
AdSystem
AppSystem
ProfileSystem
ClickAuditSystem

To make the notion of privacy accessible in the TMA system, we propose a relatively simple definition within our model: users' natural properties (e.g. age, gender and interests), which are stored in *ProfileSystem*, and users' behavioural data (e.g. browsing websites and clicking ads), which are stored in both *ProfileSystem* and

ClickAuditSystem, are at the heart of the issues of privacy with which we concern ourselves. Thus, by tracking the data flow involved in the two subsystems, there is the potential to specify how much of a user's personal information is released to others.

4.3.2 Privacy-affected operations

We now give consideration to key operations that affect privacy. For convenience, we introduce schemas that are concerned with one subsystem only and leave other subsystems unchanged. For example, with respect to *ProfileSystem*, we have *ProfileSystemOP*:

ProfileSystemOP
$\Delta System$
$\Xi ActorSystem$
$\Xi AdSystem$
$\Xi AppSystem$
$\Xi ClickAuditSystem$

This denotes the fact that the state of the overall system will change (indicated by the symbol Δ), but the states of the four named subsystems will remain unchanged (indicated by the symbol Ξ).

1. User profile maintenance.

   A user profile is a series of records created by an ad-network for a particular user that stores the user's personal data and deduced information. The maintenance process associated with user profiles is reflected in steps 6–8 of the TMA workflow of Figure 4.1. As the process takes place in the servers of the ad-networks, the users are unable to intervene in it. Therefore, the user's personal information is released to the ad-network without their control.

   The operation of creating a new user profile is defined as follows.

CreateUserProfile

$$
\begin{array}{l}
ProfileSystemOP \\
anId? : AdNetworkId \\
uId? : UserId \\
upId? : UserProfileId \\
up? : UserProfile \\
ui? : \mathbb{P}\ AdCategoryId \\
\hline
anId? \in \mathrm{dom}\ adNetwork \\
uId? \in \mathrm{dom}\ user \\
upId? \notin \mathrm{dom}\ userProfile \\
up? \notin \mathrm{ran}\ userProfile \\
\forall\, i : UserProfileId \bullet \\
\quad \neg\ (\ i \mapsto uId? \in profileOfUser \land i \mapsto anId? \notin profileInAdNetwork) \\
userProfile' = userProfile \cup \{upId? \mapsto up?\} \\
userInterest' = userInterest \cup \{upId? \mapsto ui?\} \\
profileOfUser' = profileOfUser \cup \{upId? \mapsto uId?\} \\
profileInAdNetwork' = profileInAdNetwork \cup \{upId? \mapsto anId?\}
\end{array}
$$

Users are able to specify that the ad-network *anId*? is able to identify the relationship between the user *uId*? and the profile *up*? by performing this operation. However, as the operation is performed in the servers of ad-networks instead of users' devices, the users are unable to take control of the information-releasing process.

Updating a user profile is an operation upon the *ProfileSystem* affects only a single profile. Therefore we express this operation using promotion.

The local operation schema is introduced by

UpdateUserProfileLocal

$$
\begin{array}{l}
\Delta UserProfile \\
ubi? : UserBasicInfo \\
sbi? : \mathbb{P}\ SearchBrowserInfo \\
smd? : \mathbb{P}\ SelfMadeDocument \\
\hline
userBasicInfo' = ubi? \\
searchBrowserInfo' = sbi? \\
selfMadeDocument' = smd?
\end{array}
$$

and the promotion schema is described as

PromoteUserProfile

$\Delta UserProfile$
$ProfileSystemOP$
$upId? : UserProfileId$

$upId? \in \text{dom}\, userProfile$
$\theta UserProfile = userProfile\, upId?$
$userProfile' = userProfile \oplus \{upId? \mapsto \theta UserProfile'\}$
$userInterest' = userInterest$
$profileOfUser' = profileOfUser$
$profileInAdNetwork' = profileInAdNetwork$

Thus, the global operation of updating a user profile is defined as follows.

$$UpdateUserProfile \mathrel{\widehat=} \exists \Delta UserProfile \bullet UpdateUserProfileLocal \land PromoteUserProfile$$

The operation *UpdateUserProfile* can be triggered in several conditions. For example, a user's profile might be updated after she searched with new keywords, browsed new products on online stores, or clicked on a particular ad in an app.

2. Ad selection

Ad selection is the core feature of the TMA system. Relevant ads can be selected by considering one or more factors: the user's hobbies and location; the most suitable format of ads for the active app and device; the ad budget; etc. The selection process is shown in steps 9 and 10 of the TMA workflow of Figure 4.1.

The operation of filtering ads by keyword is described via the following schema. Ad selection based on classification or brands works in a similar fashion.

ReturnAdsByKeyword

$\Xi System$
$anId? : AdNetworkId$
$k? : AdKeyword$
$a! : \mathbb{P}\ AdId$

$anId? \in \text{dom}\, adNetwork$
$a! = \{i : AdId \mid i \in (adInAdNetwork\, anId?) \land k? \in i.keyword\}$

Selecting the most relevant ads involves complex algorithms — many factors need to be considered, as do their respective weights. Here we give a simple example, which makes use of the target audience data of ads, a particular user's interests and attributes of active ad units, to pick the best ads. For the sake of readability, the weights of these factors are valued equally in the example.

SelectBestAds

$$
\begin{array}{l}
\Xi System \\
auId? : AdUnitId \\
upId? : UserProfileId \\
a! : \mathbb{P}\ AdId \\
\hline
auId? \in \mathrm{dom}\ adUnit \\
upId? \in \mathrm{dom}\ userProfile \\
adUnitInAdNetwork\ auId? = profileInAdNetwork\ upId? \\
a! = \{i : AdId \mid \\
\qquad i \in (adInAdNetwork\ (adUnitInAdNetwork\ auId?)) \land \\
\qquad (userProfile\ auId?).userBasicInfo \in (ad\ i).targetAudience \land \\
\qquad (ad\ i).format = (adUnit\ auId?).format \land \\
\qquad userInterest\ upId? \cap (ad\ i).category \neq \emptyset\}
\end{array}
$$

The operation indicates that the ad-network needs to make use of the user's basic information (e.g. age, gender, location, etc.) and interests that are stored in the user profile *upId?* in the ad selection operation.

3. Ad-clicking record

   The last steps of the TMA workflow involves recording users' clicks on ads. Since the click operations could reflect users' preferences (by assuming that users only click on ads that attract them), they can also be used as evidence for targeting and should be considered as a privacy-related feature. Again, users are unable to control the flow of their personal information within this process.

   All involved data should be referred to a single ad-network. The predicates of the following schema *AddClickAudit*, which describes the operation of adding a new record, captures this requirement. Again, the ad-network can deduce

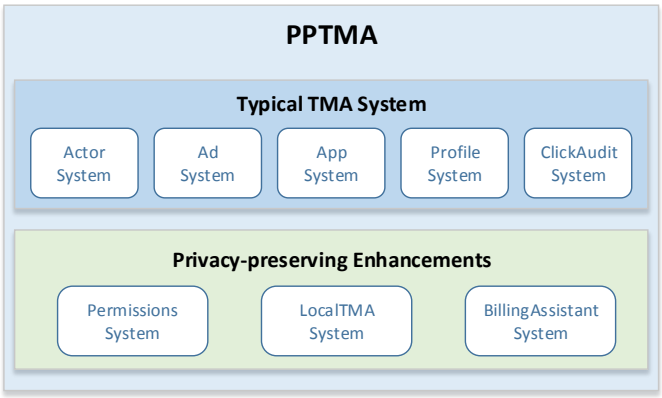

Figure 4.2: The composition of the PPTMA system

the users' interests by analysing the ad they clicked on — users' privacy is also compromised in this operation.

AddClickAudit

$$
\begin{array}{l}
ClickAuditSystemOP \\
anId? : AdNetworkId \\
cId? : ClickAuditId \\
c? : ClickAudit \\
\hline
anId? \in \mathrm{dom}\ adNetwork \\
cId? \notin \mathrm{dom}\ clickAudit \\
c? \notin \mathrm{ran}\ clickAudit \\
c?.userProfileId \mapsto anId? \in profileInAdNetwork \\
c?.adId \in adInAdNetwork\ anId? \\
c?.adUnitId \mapsto anId? \in adUnitInAdNetwork \\
clickAudit' = clickAudit \cup \{cId? \mapsto c?\} \\
clickAuditInAdNetwork' = clickAuditInAdNetwork \cup \{cId? \mapsto anId?\}
\end{array}
$$

4.4 A model of PPTMA

Having analysed the original TMA model, we now present the formal model of the PPTMA system and discuss how the formal models can help reason about issues of privacy. Figure 4.2 shows the composition of the overall system at a high level of abstraction.

4.4.1 Overview of the model

The model of the typical TMA system described in the previous subsection can help users understand how much of their personal information is disclosed. However,

in this model, users' ability to control access to their personal information is limited — they can specify the released information and involved operations, but cannot intervene in the process.

We now refine the initial model by importing a permissions mechanism, a local TMA mechanism, and a billing assistant system. This helps us to describe the core features of PPTMA. The model of PPTMA allows users to control how much of their information is released, and helps to balance privacy and utility in the ad-selection process.

In Section 4.3, we assumed that apps with ad-plugins are installed on users' mobile devices unconditionally and could obtain all resources they required for running. In that situation, apps are able to collect users' personal information and deliver ads once they were published and installed.

In the real world, however, to install an app and make it available for delivering ads on a device is much more complicated. For instance, the early versions of Android (versions before Android 6.0) employed a fine-grained permissions mechanism, whereby apps need to declare permissions for their general functions and ask for extra permissions to enable an ad-network's plugins to work properly. Upon the installation of an app, all permissions it requires are delivered to the user. Only once all permissions have been granted can the app be installed successfully.

We improve the permissions mechanism by adding a feature that is not supported in the major versions of the Android system[1]. The permissions mechanism described in this section is consistent with the access control mechanism of Android 6.0 — enabling permissions held by apps to be modified after the apps are installed. This mechanism also enables apps to work properly with corresponding permissions granted by users. The enhancement gives users the ability to control which parts of their information can be released to which apps, as well as to the related ad-networks. The subsystem *PermissionsSystem* allows us to capture this feature.

[1]At the time of writing (August 2017), the distribution number of Android 6.0 and above is 46% [84]. Thus, more than half of Android devices could not apply the new permissions mechanism officially.

PermissionsSystem

$$
\begin{array}{l}
permission : PermissionId \pfun Permission \\
installedApp : UserId \pfun \mathbb{P}\, AppId \\
permissionRequiredOfApp : AppId \pfun \mathbb{P}\, PermissionId \\
permissionRequiredOfAdNetwork : AdNetworkId \pfun \mathbb{P}\, PermissionId \\
permissionHeldOfInstalledApp : (UserId \times AppId) \pfun \mathbb{P}\, PermissionId \\
\hline
\forall\, ids1 : \mathbb{P}\, PermissionId \bullet ids1 \in \operatorname{ran} permissionRequiredOfApp \\
\qquad \Rightarrow ids1 \subseteq \operatorname{dom} permission \\
\forall\, ids2 : \mathbb{P}\, PermissionId \bullet ids2 \in \operatorname{ran} permissionRequiredOfAdNetwork \\
\qquad \Rightarrow ids2 \subseteq \operatorname{dom} permission \\
\forall\, ids3 : \mathbb{P}\, PermissionId \bullet ids3 \in \operatorname{ran} permissionHeldOfInstalledApp \\
\qquad \Rightarrow ids3 \subseteq \operatorname{dom} permission \\
\operatorname{dom} permissionHeldOfInstalledApp \in installedApp
\end{array}
$$

The function *installedApp* describes which apps have been installed on a particular user's device. Users can revoke or update permissions of these installed apps with provided operations.

The permissions required by an app for enabling it to run properly are captured by the function *permissionRequiredOfApp*, and the necessary permissions required by an ad-network to ensure its ability to collect users' data and deliver relevant ads are described by the function *permissionRequiredOfAdNetwork*. The function *permissionHeldOfInstalledApp* then tells us which permissions are held by apps that have been installed on a device.

With this subsystem, users can prevent ad-networks from collecting user data and delivering targeted ads by revoking all permissions required by related apps. This mechanism, however, compromises the ability of the advertisers and ad-networks — as their inaccurate ads might not be clicked, nor even displayed. To this end, we have implemented another extension to the model. The core mechanism creates coarse-grained copies of user profiles, pre-downloads ads to the mobile devices, then selects relevant ads from the pool of local ads according to local user profiles. The enhancement enables user profiles and targeted ads (the most significant privacy-related elements of the system) to be handled locally inside the mobile device. This mechanism is introduced by *LocalTMASystem*.

LocalTMASystem

$$customUserProfile : UserProfileId \pfun UserProfile$$
$$localAds : UserId \pfun (AdNetworkId \pfun \seq AdId)$$

The function *customUserProfile* represents different user profiles edited manually by the users. The function *localAds* describes the pre-downloaded ads inside the device. It is important to note that the custom user profile, which is maintained by the user and not accessible to the ad-networks, differs from the actual user profile. Thus, the user's personal data stored in *LocalTMASystem* will not be released to ad-networks — unless the user chooses to share the coarse-grained or fine-grained version of it. The model, therefore, helps the user to make decisions pertaining to what extent they are willing to disclose their personal information.

The final extension to the TMA model is the billing assistant system. Click-audits are obfuscated in this subsystem before being submitting to the servers of ad-networks. This feature helps to record click operations without exposing the user's information.

BillingAssistantSystem

$$obfuscatedClickAudit : ObfuscatedClickAuditId \pfun ClickAudit$$
$$clickAuditMapping : ObfuscatedClickAuditId \pfun ClickAuditId$$

$$\dom clickAuditMapping \subseteq \dom obfuscatedClickAudit$$

The function *obfuscatedClickAudit* represents the new copy of click record with an obfuscated userId. The function *clickAuditMapping* describes the association between the original click record and its obfuscated copy, which can be used to infer the original user who clicked on the particular ad when there is a necessary to verify potential click-fraud attacks.

The PPTMA model is based on the three new subsystems, together with the model of the original TMA system (again, for the sake of readability, the constraints in this schema are omitted).

PPTMA

$$System$$
$$PermissionsSystem$$
$$LocalTMASystem$$
$$BillingAssistantSystem$$

4.4.2 Improved operations

The one-subsystem-involved operations in the original TMA model (e.g. *ProfileSystemOP*) are updated by including the following definations:

$\Xi PermissionsSystem$
$\Xi LocalTMASystem$
$\Xi BillingAssistantSystem$

These new operations, *PermissionsSystemOP*, *LocalTMASystemOP*, and *BillingAssistantSystemOP*, are also added to the *PPTMA* system. Again, they work only upon the relevant subsystems.

Now we give consideration to new operations imported by PPTMA system and discuss the improvement of the original operations of the TMA system.

1. Block ads.

   The *PermissionsSystem* limits the ability of collecting user's personal information. As an extreme instance, a user with a serious privacy concern can completely stop an installed app from collecting data and delivering targeting ads by revoking all permissions required by a particular ad-network from the app.

   The operation is described as follows.

BlockAds

$$
\begin{array}{l}
PermissionsSystemOP \\
uId? : UserId \\
aId? : AppId \\
anId? : AdNetworkId \\
\hline
uId? \in \mathrm{dom}\, user \\
aId? \in installedApp\ uId? \\
\{(uId?, aId?)\} \lhd permissionHeldOfInstalledApp' = \\
\quad \{(uId?, aId?)\} \lhd permissionHeldOfInstalledApp \\
permissionHeldOfInstalledApp'\ (uId?, aId?) = \\
\quad permissionHeldOfInstalledApp\ (uId?, aId?) \backslash \\
\qquad permissionRequiredOfAdNetwork\ anId? \\
permission' = permission \\
installedApp' = installedApp \\
permissionRequiredOfApp' = permissionRequiredOfApp \\
permissionRequiredOfAdNetwork' = \\
\quad permissionRequiredOfAdNetwork
\end{array}
$$

2. Update permissions. Apart from the extreme case, the other users who want to preserve particular parts of their personal information can achieve the goal by limiting related permissions. For example, revoking the location-related permissions will help to hide the user's current location, and revoking SMS-related permissions prevents advertisers from collecting keywords in the self made messages. The operation of updating permissions is described as follows.

$$
\begin{array}{|l}
\textit{UpdatePermissions} \\
\hline
\textit{PermissionsSystemOP} \\
\textit{uId}? : \textit{UserId} \\
\textit{aId}? : \textit{AppId} \\
p? : \mathbb{P}\, \textit{PermissionId} \\
\hline
\textit{uId}? \in \operatorname{dom} \textit{user} \\
\textit{aId}? \in \textit{installedApp}\, \textit{uId}? \\
\forall\, i : \textit{PermissionId} \bullet i \in p? \Rightarrow i \in \operatorname{dom} \textit{permission} \\
\{(\textit{uId}?, \textit{aId}?)\} \lhd \textit{permissionHeldOfInstalledApp}' = \\
\quad \{(\textit{uId}?, \textit{aId}?)\} \lhd \textit{permissionHeldOfInstalledApp} \\
\textit{permissionHeldOfInstalledApp}'\, (\textit{uId}?, \textit{aId}?) = p? \\
\textit{permission}' = \textit{permission} \\
\textit{installedApp}' = \textit{installedApp} \\
\textit{permissionRequiredOfApp}' = \textit{permissionRequiredOfApp} \\
\textit{permissionRequiredOfAdNetwork}' = \\
\quad \textit{permissionRequiredOfAdNetwork} \\
\hline
\end{array}
$$

3. Pre-download ads.

Before we can select relevant ads inside the device, a number of ads need to be pre-downloaded from the servers of ad-networks to devices. To improve the click rate later, we provide interests information abstracted from the user's coarse-grained profile to the server to narrow down the categories of pre-downloaded ads. Since we submit the interests without a user identifier, the server is only aware that an anonymous user is asking for a set of ads. Therefore users' privacy will not be compromised during this process.

AddLocalAd

$$
\begin{array}{l}
LocalTMASystemOP \\
uId? : UserId \\
anId? : AdNetworkId \\
ads? : \mathrm{seq}\ AdId \\
\hline
uId? \in \mathrm{dom}\ user \\
anId? \in \mathrm{dom}\ adNetwork \\
\mathrm{ran}\ ads? \subseteq adInAdNetwork\ anId? \\
\{uId?\} \lhd localAds' = \{uId?\} \lhd localAds \\
localAds'\ uId? = localAds\ uId? \oplus \{anId? \mapsto ads?\} \\
customUserProfile' = customUserProfile
\end{array}
$$

4. Local user profile-maintenance, ad selection, and click-audit.

   With the local TMA system, we can create and maintain custom copies of user profiles inside the device. Therefore users' personal information is stored in their own devices instead of being submitted to ad-networks' servers. As a result, users' personal information will not leak out to the ad-networks.

   It follows that ad-selection operations are composed of two stages. The first stage involves selecting and pre-downloading potential ads on remote servers with respect to coarse-grained copies of user profiles. In the second stage the most relevant ads are selected from the pre-downloaded ads by analysing the fine-grained user profiles on a local client. These operations can be implemented (in both stages) via custom algorithms.

   The operations of maintaining local user profiles and selecting ads from the local cache in the PPTMA model are quite similar to those corresponding operations in the TMA model. Therefore, we will not repeatedly present the operations here. It is worth noting that, in the PPTMA model, users are able to intervene in these operations as they are performed locally on users' devices.

   The click-audit is obfuscated in *BillingAssistantSystem* before being submitted to an ad-network. The original *UserId* value is replaced with a random single-use identifier to ensure that the ad-network cannot identify the specific

user. The mappings of the original and the obfuscated click-audits are maintained in the *BillingAssistantSystem* to enable the tracing of click-fraud attacks. The obfuscated click-audit is sent to an ad-network from the *BillingAssistantSystem* rather than the mobile users. Therefore, the meta-information of the connection cannot be used to identify the original users.

The related operation is described as follows.

ObfuscateClickAudit

$BillingAssistantSystemOP$
$cId? : ClickAuditId$
$ocId? : ObfuscatedClickAuditId$
$randomId? : UserId$
$c? : ClickAudit$

$cId? \in \text{dom } clickAudit$
$ocId? \notin \text{dom } obfuscatedClickAudit$
$ocId? \notin \text{dom } clickAuditMapping$
$cId? \notin \text{ran } clickAuditMapping$
$randomId? \notin \text{dom } user$
$\forall ca : ClickAudit \bullet ca \in \text{ran } clickAudit \Rightarrow ca.userId \neq randomId?$
$c?.userId = randomId?$
$c?.adId = (clickAudit\ cId?).adId$
$c?.adUnitId = (clickAudit\ cId?).adUnitId$
$c?.date = (clickAudit\ cId?).date$
$obfuscatedClickAudit' = obfuscatedClickAudit \cup \{ocId? \mapsto c?\}$
$clickAuditMapping' = clickAuditMapping \cup \{ocId? \mapsto cId?\}$

4.5 Application of the PPTMA model

Having described the PPTMA model, we now present audience targeting as an instance to show how a mainstream ad-selection mechanism can be applied in a privacy-friendly way with our models. For the sake of brevity, we discuss only one instance of several different tests of the models. The instance illustrates how (the implementations of) these models can assist users in controlling how much of their personal information should be released to the ad-network, and help them to specify which particular operations disclose corresponding information.

4.5.1 The first stage of the ad-selection process: pre-download ads

By analysing a user's profile, ad-networks can assign the user to a particular audience segment, then recommend relevant ads for the user. The segment indicates the basic information and interests of associated users.

AudienceSegment

$$userBasicInfo : UserBasicInfo$$
$$interestKeywords : \mathbb{P}\, Keyword$$

We introduce one type explicitly — *Age* — to demonstrate the role that formal models can play in obfuscation.

$$Age ::= actual\langle\langle \mathbb{N} \rangle\rangle \mid range\langle\langle \mathbb{N} \times \mathbb{N} \rangle\rangle$$

Here, an age can either be a specific age, or drawn from a range.

We assume that there is a user whose basic information is described as follows.

$$UserBasicInfo1 = \langle\!|\ gender == Male, age == actual(25), \\ location == Oxford, language == English \ |\!\rangle$$

We assume that the profile ID of this user is $UserProfileId1$. The user is interested in *Basketball* (which we assume has the associated identifier *IdForBasketball*); therefore, in *ProfileSystem* the following predicate holds.

$$IdForBasketball \in userInterest(UserProfileId1)$$

The first stage of ad selection then consists of the following processes.

1. *Generate coarse-grained copy of the user's profile.*

   The schema *AudienceSegment* suggests that the user's basic information $UserBasicInfo1$ and interest *IdForBasketball* might be released in the following operations. The user chooses to only submit coarse-grained information to the ad-network, rather than his precise profile. Therefore he generates a custom user profile with the following basic information:

   $$UserBasicInfo2 = \langle\!|\ gender == DeclineToState, age == range(20, 30), \\ location == UK, language == English \ |\!\rangle$$

The custom profile is associated with *UserProfileId*2. Instead of disclosing his interest of Basketball, he only share his interests at a higher level as *Team Sports*. Therefore we have:

$$IdForTeamSports \in userInterest(UserProfileId2)$$

2. *Assign the user to a relevant audience segment.*

Based on the submitted profile — *UserProfileId*2 — the user will be assigned to the audience segment *AudienceSegment*2. By contrast, the original profile *UserProfileId*1 will lead the user to *AudienceSegment*1.

$$\begin{aligned} AudienceSegment1 = \langle\!| \ & userBasicInfo == UserBasicInfo1, \\ & interestKeywords == \{Basketball\} \ |\!\rangle \\ AudienceSegment2 = \langle\!| \ & userBasicInfo == UserBasicInfo2, \\ & interestKeywords == \{TeamSports\} \ |\!\rangle \end{aligned}$$

By analysing the two copies of audience segments, the user can understand which parts of his personal information is released (and to what extent).

3. *Select potential ads for the user.*

A set of potential ads related to the segment can be selected via the following operation.

SelectAdsByAudienceSegment

$$\begin{aligned} & \Xi PPTMA \\ & as? : AudienceSegment \\ & anId? : AdNetworkId \\ & ads! : \mathbb{P}\, AdId \\ \hline & ads! = \{i : AdId \mid \\ & \quad i \in (adInAdNetwork\ anId?) \land \\ & \quad as?.userBasicInfo \in (ad\ i).targetAudience \land \\ & \quad as?.interestKeywords \cap (ad\ i).keyword \neq \emptyset\} \end{aligned}$$

Here, the ad-network applies *AudienceSegment*2, which is abstracted from the coarse-grained user profile, as the input *as*?. Therefore, ads associated with *TeamSports* (e.g. Football, Basketball, Baseball, Handball, etc.) will be selected. In addition, these ads are all applicable to a person who is aged 20 to 30, lives in the UK, and speaks English.

4. *Rank and deliver ads.*

   The selected ads are ranked on the servers without disclosing particular ranking strategies (e.g. ads can be sorted by remaining ad budgets, publication date, distance from the current location, etc.) that are applied by different ad-networks. The ordered list is then pre-downloaded to the user's device.

4.5.2 The second stage of the ad-selection process: local ad selection

Assuming that, via the first stage of ad selection, the user has obtained 100 ads related to different team sports located in different places in the UK, the local ad-selection stage can then help to pick the most relevant ads according to the user's precise profile. The processes are described as follows.

1. *Generate the precise audience segment from the fine-grained user profile.*

   As discussed in Section 4.5.1, *AudienceSegment*1, which is more precise than *AudienceSegment*2, can be abstracted from the original user profile associated with *UserProfileId*1. Since *UserProfileId*1 and *AudienceSegment*1 are both maintained locally in the user's mobile device, no personal information is released in this process.

2. *Select the most relevant ads.*

   With the precise audience segment, less relevant ads can be filtered out from the list of potential ads. For example, since we know the user's precise interest is Basketball, ads associated with Football, Baseball and Handball can all be removed from the list. In the same way, ads based in the UK, but outside of Oxford can also be filtered out. Note that the formats of selected ads should be consistent with the ad units of the active app.

$$
\begin{array}{|l}
\textit{SelectMostRelevantAds} \\
\hline
\Xi PPTMA \\
uId? : UserId \\
as? : AudienceSegment \\
auId? : AdUnitId \\
adsSet! : \mathbb{P}\, AdId \\
adsList! : \text{seq}\, AdId \\
\hline
uId? \in \text{dom}\, localAds \land auId? \in \text{dom}\, adUnit \\
adsSet! = \{\ i : AdId \mid \\
\qquad (i \in \text{ran}((localAds\, uId?)\, ((adUnit\, auId?).adNetwork)) \\
\qquad \land \\
\qquad as?.userBasicInfo \in (ad\, i).targetAudience \\
\qquad \land \\
\qquad as?.interestKeywords \cap (ad\, i).keyword \neq \emptyset \\
\qquad \land \\
\qquad (ad\, i).format = (adUnit\, auId?).format)\} \\
adsList! = \\
\quad ((localAds\, uId?)\, ((adUnit\, auId?).adNetwork)) \upharpoonright adsSet! \\
\hline
\end{array}
$$

Finally, we obtain a shortlist of ads with their relative ranks decided by the ad-network. The top ads on the list can then be displayed in apps as the most relevant ads. The two-stage ad-selection process helps to balance privacy and utility: ad-networks can only obtain the coarse-grained information that users would like to disclose, and users are able to obtain the most relevant ads based on their fine-grained profile. Importantly, the involved data and operations are easy for users to specify with the help of the formal models.

4.5.3 Click-audit obfuscating and click-fraud detecting

Finally, the user clicks on the displayed ad, and a click-audit record is created. As opposed to the second stage of the ad-selection process, the click operation and audit should be submitted to the ad-network, rather than stored in the mobile device. Thus, the user's interest might be deduced by analysing the clicked ad.

In order to prevent information leakage, the click-audit needs to be processed before being delivered to the ad-network. The click-audit obfuscating and click-fraud detecting mechanisms are described as follows.

1. *Obfuscate user identifier for an ad click report.*

   As discussed in Section 4.4, a random user identifier, $RandomId1$, is generated in the billing assistant system to replace the original user identifier, $UserId1$. $ClickAudit2$, the obfuscated copy of $ClickAudit1$, will then be submitted to the server of related ad-network. The mappings of the two copies are stored in the subsystem for later use.

$$
\begin{aligned}
&ClickAudit1 = \langle\!| \ userId == UserId1, adId == AdId1, \\
&\qquad\qquad adUnitId == AdUnitId1, date == Date1 \ |\!\rangle \\
&ClickAudit2 = \langle\!| \ userId == RandomId1, adId == AdId1, \\
&\qquad\qquad adUnitId == AdUnitId1, date == Date1 \ |\!\rangle \\
&BillingAssistantSystem = \\
&\quad \langle\!| \ obfuscatedClickAudit == \\
&\qquad\qquad \{ObfuscatedClickAuditId1 \mapsto ClickAudit2\}, \\
&\quad\ clickAuditMapping == \\
&\qquad\qquad \{ObfuscatedClickAuditId1 \mapsto ClickAuditId1\} \ |\!\rangle
\end{aligned}
$$

2. *Detect click-fraud attacks.*

   The feature of click-audit obfuscating will not affect original click-fraud detecting mechanisms applied by ad-networks. As an example, bait ads [24, 28] are hardly clicked by humans, but are regularly clicked by automated bots. For example, the content of an ad is completely related to *Football*, but all attributes hidden behind the ad might be assigned to *Basketball*. A human user who is interested in Basketball might deem this ad a failed recommendation and ignore it. On the other hand, a bot performing click-fraud will be more likely to click on the ad without realising the inconsistent content. Thus, the ad-network can use click-audits of bait ads to trace suspected malicious users.

   Given an obfuscated click-audit of a bait ad, the real user can be identified with the permission from *BillingAssistantSystem*.

Involved operations	User-held information example	Released information example	AdNetwork-held information example	Effects
Pre-download operations	*UserId*1 *Male* 25 *Oxford* *English* *Basketball*	1. Obfuscation: Age, Location, Interest 2. Disclosure: Language	*Null* *Null* $20-30$ *UK* *English* *TeamSports*	1. Ad-networks obtain the coarse-grained data of *someone* who cannot be identified. 2. Related ads are selected for the *someone*.
Local ad selection operations	As above	No data is released	As above	1. The precise information is well preserved. 2.The most relevant ads can be selected.
Click-audit operations	As above, and: *UserId*1 *AdId*1 *AdUnitId*1 *Date*1	1. Obfuscation: UserID 2. Disclosure: ClickedAd, AdUnit, Date	As above, and: *RandomId*1 *AdId*1 *AdUnitId*1 *Date*1	Ad-networks cannot deduce the original user's interests by analysing click-audits.

Table 4.1: Analysis on released personal information and related effects: an example

ClickFraudDetect

$$\Xi PPTMA$$
$$ocId? : \mathbb{P}\ ObfuscatedClickAuditId$$
$$uId! : \mathbb{P}\ UserId$$

$$ocId? \subseteq \text{dom}\ obfuscatedClickAudit$$
$$uId! = \{u : UserId \mid (\forall o : ObfuscatedClickAuditId \bullet u = (clickAudit\ (clickAuditMapping\ o)).userId)\}$$

4.6 Analysis

The main purpose of our model is to make the control of personal information easily specified by users and enable users to understand the effects of their decisions,

rather than measuring the privacy leakage. Hence, four basic metrics from a user's perspective are used to compare the different impacts introduced by our models, which are:

1. The amount and granularity of lost ownership and control following an operation.

2. The amount and granularity of information shared initially by users themselves following an operation.

3. The amount and granularity of leaked information following an operation.

4. The probability that a user's original interests can be deduced by a single third-party.

We have used ProZ [118] to analyse our model. ProZ allows its users to control the order in which operations are performed after the model is initialised. It also provides the ability to animate randomly.

We first performed operations involved in the TMA workflow, then animated new features associated with PPTMA. The result suggests new features merge well with the original TMA system and gives confidence in our prototype solution. The result suggests that our model contains the necessary state and operations to describe the basic workflow of the current TMA system. Then we animated the model by performing new operations imported by *PermissionsSystem*, *LocalTMASystem*, and *BillingAssistantSystem*. The animation suggests that new operations merge well with those from the original TMA model.

We also animated the model via a sequence of 500 randomly chosen operations. The operation history list suggests that operations execute in a logical order due to the well-defined constraints. For example, an ad can be published only after its owner — the advertiser — has been added into the system; best ads can be selected only after ads and apps have been published, user profiles have been created, and apps have imported ad-plugins from the same ad-networks related to those ads.

We then performed model-checking with the mixed breadth-first and depth-first search strategies. No deadlocks or invariant violations were detected from the model in more than 2,000,000 distinct states and 4,000,000 transitions.

We paid particular attention to our main focus, which is how these models (and the related implementations) might help users to control how much of their personal information is released to the ad-network, to specify which particular operations release corresponding information, and to understand how their control might affect the ad-selection and user-tracking processes. Table 4.1 illustrates this and discusses the above mentioned four basic metrics. The analysis is based on the instance described in Section 4.5. Furthermore, all states and operations can be traced back by checking the state properties and the operation history list. Therefore, we can identify the source of each ad, ad unit, app and user profile involved in the process, which, in turn, provides the ability for us to detect malicious operations such as click-fraud attacks.

4.7 Summary

In this chapter, we have shown how formal models might be used in helping to reason about the balance between benefits of mobile users and advertising corporations in the context of TMA. In particular, we have shown, in the spirit of Tschantz and Wing's contribution [115], the beneficial roles that formal models can play in reasoning about privacy. In our specific context, formal models allow users to specify the control of their personal information, and help them to understand how this control would affect the processes of ad selection and user tracking.

Based on requirements established in the previous chapter and the formal models described in this chapter, we will discuss the design of the *Privacy-Preserving Targeted Mobile Advertising (PPTMA)* framework in the following chapter.

5

PPTMA: A framework for privacy-preserving targeted mobile advertising

Contents

5.1 Design guidelines 102
5.2 System architecture 104
5.3 Prototype implementation 107
5.3.1 Ad scanning 107
5.3.2 Permissions and personal profile management 110
5.3.3 Sandbox 113
5.3.4 Simulation and evaluation 115
5.3.5 Privacy improvements introduced by PPTMA 118
5.4 Summary 120
5.4.1 Contribution over previous solutions 121
5.4.2 Possible objections 122

In the previous chapter, we discussed how formal models can help in underpinning the prototype PPTMA system, in analysing privacy in the context of targeted mobile advertising, and in allowing users to specify control of their personal information. Based on the formal models, this chapter is concerned the design of the prototype system. We start by discussing the design guidelines that are adopted in our solution.

The contents of this chapter are based on *Yang Liu and A. C. Simpson: Privacy-*

preserving targeted mobile advertising: Requirements, design, and a prototype implementation. In: Software: Practice and Experience 46.12 (2016) 1657 — 1684 [11].

5.1 Design guidelines

Haddadi *et al.* [120] describe the concept of a *databox* — a trusted platform with facilities for data management, a fine-grained access control mechanism for personal information, means for users to interact with their own data, and the ability to support innovative uses with commercial organisations. In many ways, our system is an instantiation of such a databox. In addition, we have adopted the foundational principles of *Privacy by Design* (PbD) [121], which serves as a reference framework that guides the system design with detailed criteria for privacy preservation. Altogether, the following are taken into account in our design process.

1. **Meet the requirements of a databox.** The architecture should be designed to meet key elements of a databox: not only should it provide centralised management of user privacy, but it should also enable users to interact with their profile and support legitimate uses for cooperative ad-networks.

2. **Compliance with the PbD principles.** The maximum degree of privacy should be set as default, and our design should pursue the objective in a positive-sum win-win manner. The design attempts to balance the relationship between the needs of ad-networks and the needs of mobile users — while we respect user privacy, our design explores a manner to achieve a positive-sum approach instead of a zero-sum one.

3. **Ease-of-use.** Our design should be simple enough and straightforward for users. It may be argued that, for those users who do not have significant concerns about privacy issues, a complicated system might reduce the likelihood that they accept the approach.

In addition, a number of non-functional requirements also come to mind in the early design stage as supplements to the particular guidelines.

1. **Energy and bandwidth cost.** The study conducted by Vallina-Rodriguez *et al.* [122] suggests that network traffic and energy consumption caused by mobile advertising can be a significant fraction of the total cost of mobile users. Similar to some advertising services, our system must keep running in the background on mobile devices; as such, the energy consumption level should be reasonable. The system is positioned between the database containing sensitive information and third-party apps to prevent them from interacting directly, so there will be additional communication time and network bandwidth consumption; the relevant cost to mobile devices should also be kept as low as possible. In addition, the performance of behavioural targeted algorithms should be optimised.

2. **Effectiveness and security of data management.** Mobile devices carry much more personal information than PCs do, and this information should be handled appropriately. The inherent complexity of the privacy-related data inside mobile devices makes it difficult to manage all kinds of different information effectively. Clearly, if we cannot assure the security of our system, it might be used by malicious parties to collect users' personal information.

3. **Accuracy of advertisement selections.** Since our objective is to develop a system that helps both consumers and advertisers, the accuracy of those displayed ads needs to be ensured. For that purpose, we need to propose effective targeting strategies to select more accurate ads for the cooperative ad-networks. (Inevitably, the accuracy will be reduced, but we aim to ensure that this reduction in accuracy is tolerable.)

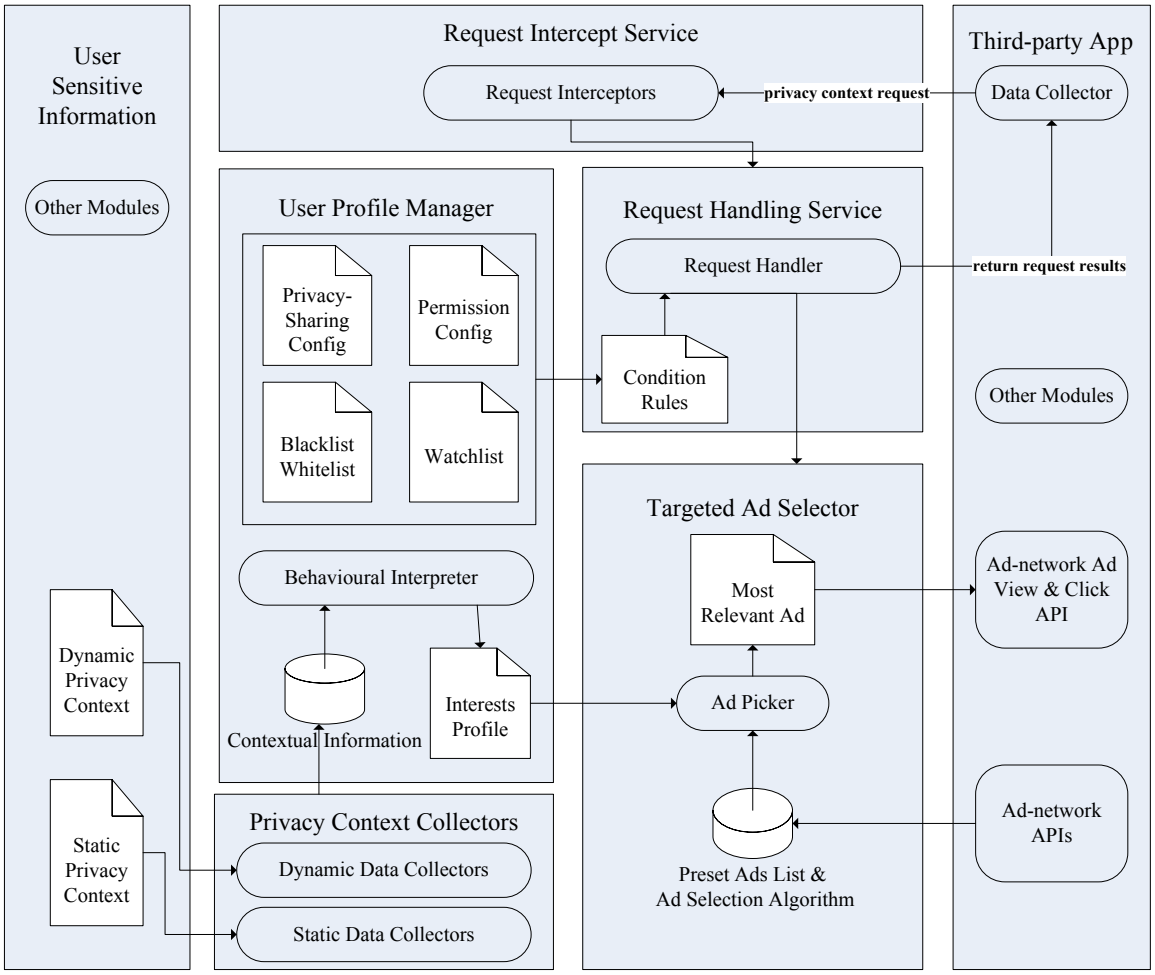

Figure 5.1: PPTMA architecture

5.2 System architecture

At a high level, our system is implemented as a service running in the background of the mobile system with an associated application to adjust the privacy settings. The system works as a middleware instantiation on mobile platforms, positioned between the underlying database and untrusted third-party mobile applications.

Figure 5.1 shows the system architecture of our prototype system. There are five main components in our design: the request intercept service; the request handling service; privacy context collectors; the user profile manager; and the targeted ad selector. We consider each in turn.

1. **Request intercept service.** The request intercept service is a background service, which prevents third-party apps from collecting different privacy-

related information, such as location, and from accessing the call log directly. Each time a third-party app asks for the user's personal information by calling related system APIs, the corresponding interceptor will block the request by hooking those APIs. The details of the request — such as the target privacy context, the request time, and the name and version of the third-party app — will then be delivered to the request handler.

2. **Request handling service.** By analysing details of requests and checking privacy-sharing settings, the request handler computes appropriate request results and returns them to the third-party apps. For example, if the information of the third-party app indicates that it contains the plugins of a cooperative ad-network, the request will be delivered to the targeted ad selector to provide the most relevant ad locally. If the third-party app is on the user's whitelist, the request handler will cancel the block and enable the app to call related APIs directly. If fine-grained access control of the given app was set by the user, the request handler will grant corresponding permissions by checking related settings from the profile manager.

3. **Privacy context collectors.** There are various privacy context collectors that can be classified as dynamic or static data collectors. These collectors monitor and fetch all privacy contexts (see Section 2.2.2) to update the contextual information database for the centralised maintenance. For example, the location information is collected by calling related APIs every 30 minutes. All location records are then saved into the contextual information database.

4. **User profile manager.** The user profile manager consists of the following parts:

    - the contextual information database, which stores various personal information collected from the mobile device;
    - the behavioural interpreter, which computes the data from the contextual information database to construct the user interest profile;

- the user interest profiles, which can be used to infer the user's interests;
- the privacy-sharing config that describes the overall privacy sharing level;
- the permission config that describes the fine-grained access control for each app;
- a blacklist and a whitelist that explicitly indicate the threat level of apps; and
- a watchlist that records those apps that need to run in a sandbox.

The user profile is not universal across all cooperative ad-networks. Set by the user, the profile could be instantiated to different copies for each ad-network. Each copy consists of a configuration file and a set of data, either authentic or fake, that record the user's basic information (e.g. location, operations on apps, hardware, installed apps, and so on) and deduced interests. By default, the original user profile will be shared, inside the mobile device, with ad-networks that appear on the whitelist. In addition, the related files and data will be obfuscated or encrypted to protect the user profile from malicious apps.

5. **Targeted ad selector.** The targeted ad selector pre-downloads some ad lists from cooperative ad-networks and stores the ad selection algorithm that they share. When combined with the interest profile from the user profile manager, the ad selector is then able to select the most relevant ad to display.

In addition, our system imports the limited versions of the ad-SDKs of cooperative ad-networks to perform the necessary functions — pre-downloading ads by providing only limited anonymous user information and submitting view and click reports without particular user identifiers. Thus, cooperative ad-networks and related apps will have no need to call the APIs of our system.

5.3 Prototype implementation

The prototype implementation of our architecture is an Android service associated with an app. The service starts automatically after the booting of the system for the purpose of hooking sensitive APIs and monitoring malicious behaviour, while the associated app is used to adjust the related privacy configuration. The prototype consists of five functional modules:

1. The **ad scanning** module scans plugins of ad-networks from installed apps and provides support for cooperative ad-networks.

2. The **permissions management** module works as a complement to the existing permission mechanism and offers fine-grained access control over each installed app.

3. The **personal profile management** module enables users to maintain their personal information and create mock copies.

4. The **sandbox** module monitors sensitive operations performed by untrusted apps.

5. The **whitelist and blacklist** module divides installed apps to different groups — trusted or malicious apps.

5.3.1 Ad scanning

Users can determine whether or not an app contains ads and then configure it for further processing. The method we use to scan ads is feature library comparing. To collect personal information and deliver ads with apps, developers need to register their apps with ad-networks, import their libraries, declare related permissions, and call ad-APIs provided by ad-SDKs.

The typical process of ad scanning can be described briefly as follows.

1. Call related Android APIs to get the source locations of all installed apps.

Table 5.1: Different ad styles supported by *Domob*

Classname in domob SDK	Represented ad style
`cn.domob.android.ads.AdView`	Banner ads
`cn.domob.android.ads.FeedsAdView`	Feeds ads
`cn.domob.android.ads.InterstitialAd`	Interstitial ads
`cn.domob.android.ads.RTSplashAd`	Real time splash ads
`cn.domob.android.ads.SplashAd`	Splash ads

2. Get the `.dex` files (Dalvik Executable files, converted from Java `.class` files) from the source locations.

3. Traverse all classes of `.dex` files, and compare the classnames with those related to ad libraries.

If the particular classnames are found in the `.dex` files, then we can determine that the involved apps do contain ads provided by the particular ad-networks. In addition, as the permission of getting source locations of apps is allowed by default, our framework does not need any extra permissions.

For instance, a typical app with ad plugins provided by the ad-network `Domob` may work as follows:

1. Register with `Domob` and get the unique publisher ID for the app.

2. Import the ad library `domob_android_sdk.jar` to the app, then declare requested permissions such as `READ_PHONE_STATE`, `WRITE_EXTERNAL_STORAGE`, `ACCESS_COARSE_LOCATION`, etc.

3. Call related ad-APIs according to different styles of ads to be displayed (see Table 5.1).

Ad libraries are usually compiled to Java Archive (`.jar`) files, which can be detected from an installed app or its installation package. By comparing jars, classnames and used APIs in apps with typical ad libraries, we can obtain the following information of a particular app (see Figure 5.2):

1. The kinds of ad styles an app contains, such as banner ads or pop-up ads, which can be deduced from the called ad-APIs and related classnames (see Table 5.1).

2. The list of ad behaviour deduced from the permissions declared in the app, such as collecting device ID or location.

3. The particular ad-networks an app contains, obtained by scanning the imported jars.

After the scanning, users can adjust the privacy strategy of involved apps to protect their personal information. Options include:

1. Uninstall the app.

2. Block ads. Users can withdraw permissions required by the ad-plugins to block ads. This strategy can also be executed by decompiling the Android Application Package (APK) or blocking related APIs.

3. Mock the information. Users can provide the preset information, which can be mocked in the personal profile management module, instead of their real information to ad-networks.

4. Provide coarse-grained information. In this strategy, only some coarse-grained information will be shared with ad-networks, such as the user's approximate location and gender.

5. Prompt each time. The system will ask the user to make a decision explicitly each time the app asks for a certain kind of personal information (see the last user interface in Figure 5.2).

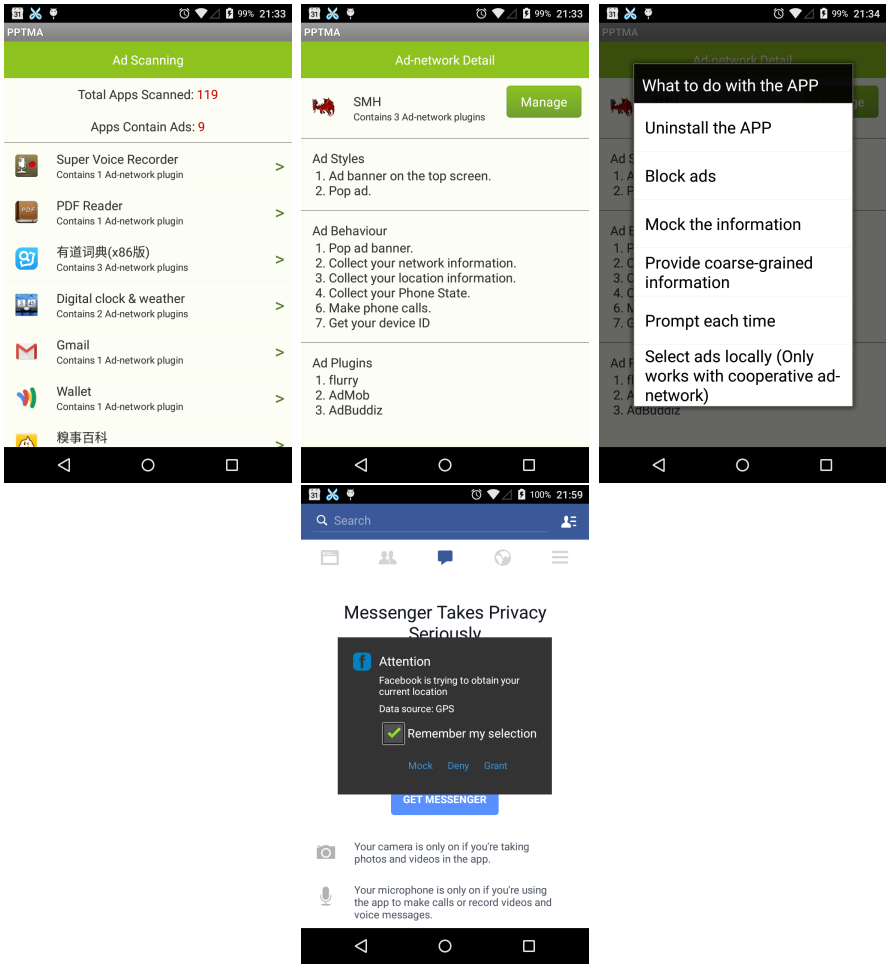

Figure 5.2: Ad scanning module

6. Select ads locally. For apps containing cooperative ad-networks, the system can run an ad selection algorithm locally.

5.3.2 Permissions and personal profile management

To collect more personal information for TMA, many apps require extra permissions that they do not actually need for general app functions. In the permissions management module, users can browse how many permissions each app holds and revise them. By doing so, the take-it-or-leave-it approach of the current Android permissions mechanism can be addressed — users can grant all permissions for the purpose of installing an app, then draw back permissions to protect personal information.

Typically, there are three approaches to implementing such an extension: app decompiling, firmware modifying, and API hooking. We discuss each in turn.

App decompiling. Android apps are actually archive files in zip format with the filename extension `.apk`. Each APK file contains the file `AndroidManifest.xml`, which declares all permissions the app requires, and the file `classes.dex`, which holds all of the app's code. By decompiling and repacking the APK file, we can remove certain permissions from the `AndroidManifest.xml` file or insert stubs around sensitive APIs, which can be detected in the `classes.dex` file. This approach can be applied to all Android systems and devices without root permission. As an app-based approach, it enables users to delete any Android permissions an app requires without damaging the system. However, there are still disadvantages with this approach. For example, after the decompilation, users need to reinstall the app — it is inevitable that some configuration and history may be lost in this process. In addition, some pre-installed apps cannot be uninstalled or reinstalled. The digital signature of the repackaged app will be modified, so those apps with extra self-checking mechanisms may not work after the repackaging. Forcibly deleting some permissions may cause 'force close' problems when the app performs operations without the related permissions declared in the `AndroidManifest.xml` file. Moreover, the modifying of APK files may even cause copyright issues.

Firmware modifying. To implement permission extensions, some researchers and developers provide patches for the existing Android system. An alternative approach involves taking advantage of the open nature of Android and modifying the source code, with a view to leveraging third-party firmware. This approach offers many benefits and allows developers to make any changes to the permissions system. However, the system patching or the firmware-flashing process relatively complicated and unreliable: users without the relevant knowledge and skills can easily damage their system. In addition, system patching only matches particular Android systems, while firmware flashing only matches particular Android devices.

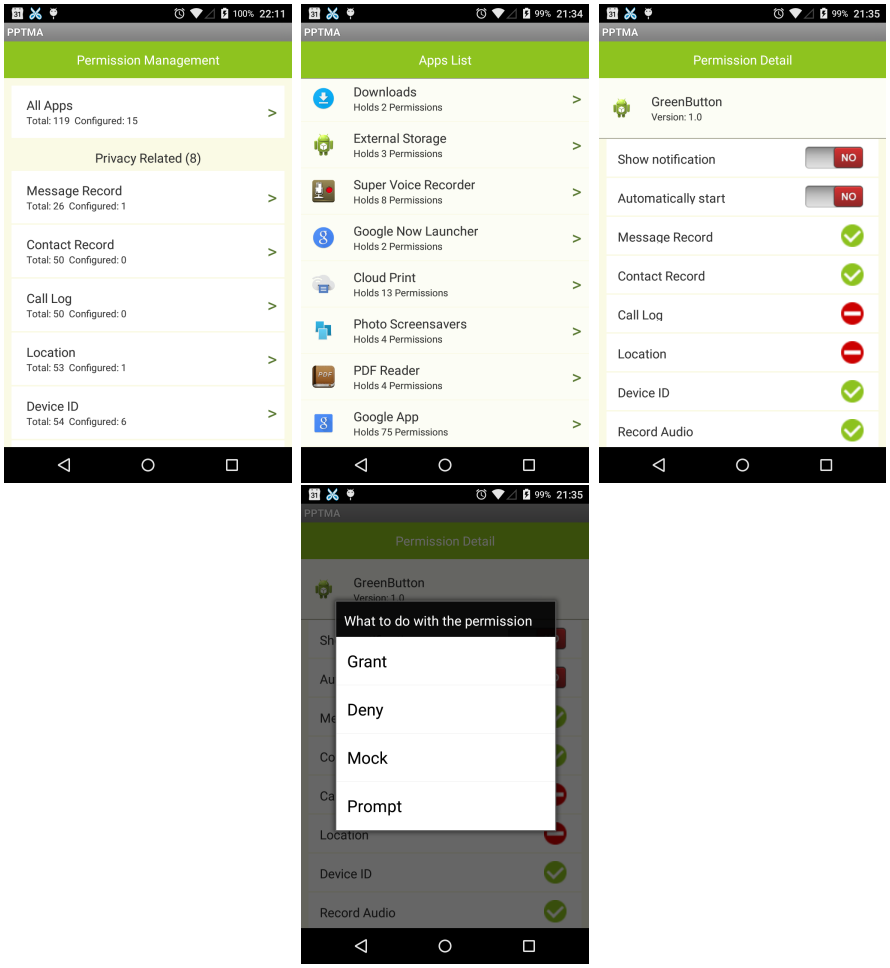

Figure 5.3: Permissions management module

Therefore the applicability and ease-of-use of this approach means that it does not meet our design guidelines.

API hooking. Apps collect users' personal information and execute permissions primarily by calling Android APIs. By hooking related APIs we can block their requests, withdraw certain permissions, and modify the return value of APIs with fake information. There are some shortcomings with this approach: it can only be applied on devices with root permission, and it requires a service to run in the background continuously to monitor callings of APIs, which may increase power consumption and impact the performance of installed apps.

Taking the above into account, as well as factors such as portability and ease-of use, we have implemented our prototype via the API hooking approach.

Much of the Android framework is implemented in Java, with a significant amount of code compiled to native machine code. Hence, we can execute the code injection and hook the system in two layers — the Java layer and the native layer.

While hooking in the native layer could offer more powerful functions, most apps collect personal information — location, installed apps, etc. — by calling related Java APIs. Therefore, for the purpose of detecting callings of such APIs, providing return values with mock data, and running our own ads selection algorithm instead of those of the ad-networks, we mainly hook sensitive Java APIs at run-time. In the current version of our prototype, we use Cydia Substrate [123], a platform for customising software on Android, to accomplish API hooking.

The user interface of this module is shown in Figure 5.3. For each permission, there are four options: grant, deny, mock, and prompt.

In the Android system, denying a permission completely may sometimes result in a 'force close', therefore the prototype does not 'deny' permissions, but, rather, returns default fake values when apps call related APIs — for example, returning 'latitude = 1.0 and longitude = 1.0' for location, or 'phoneNumber = 00000000000' for the phone state.

In the personal profile management module (see Figure 5.4), users can preset different copies of their personal profiles, whether real (collected by privacy context collectors of PPTMA) or fake (edited by users themselves). The selected edition will be provided to the third-party apps when mocked information is required in the permissions management module or the ad scanning module. For example, a user living in Oxford can make fake copies to offer a different location: she can decide how accurate the information that she would like to share with third-party apps should be.

5.3.3 Sandbox

Users can add an unknown app, an app that may contain dangerous or malicious ads, or an app that may collect their personal information into the sandbox. The

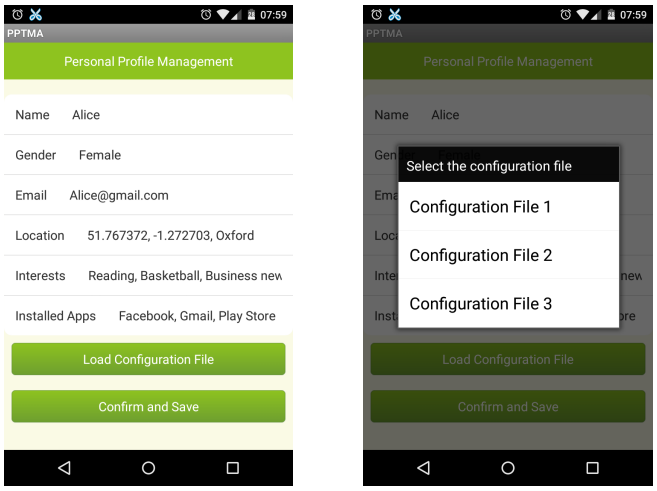

Figure 5.4: Personal profile management module

system will monitor incoming and outgoing network data, information leaks via the network, and suspect operations performed by the app. For instance, when a malicious app in the sandbox is trying to obtain the user's location, it needs to call the `android.location.LocationManager.requestLocationUpdates` API to update the user's most recent location, and it also needs to call `android.location.LocationManager.getLastKnownLocation` to get the particular information. By hooking relevant APIs, we can monitor the app's operations related to location information.

By monitoring calls of the `android.telephony.SmsManager.sendTextMessage` API, we can observe whether an app inside the sandbox is trying to send an SMS message without the user's knowledge. In addition, we can also obtain the destination, the source, and the text of the SMS message. According to different preset privacy-preserving levels, the sandbox can not only monitor the behaviour of malicious apps, but can also block their operations or even modify parameters and results involved in those operations.

The Android OS is based on the Linux kernel, and system calls of the Linux kernel can provide useful functions to apps when they perform operations on network or files. Accurate information about behaviours of apps can be obtained by capturing and analysing the system calls [124]. Other than system calls, file system logs and APIs provided by the Android system (e.g. `Android.net.TrafficStats`,

Table 5.2: Software testing environment

Type of apps	Quantity	Representative
Free off-the-shelf apps	60	Facebook, Doodle Jump
System built-in apps	69	Gmail, Package Access Helper
Apps for daily use	59	BBC iPlayer Radio, OfficeSuite Pro 8
Dummy apps	10	DummyBanner, DummyInterstitial
PPTMA client	1	PPTMA
API hooking auxiliary	1	Cydia Substrate

which can be used to monitor the network traffic of each app) can also be used to constitute the sandbox mechanism.

Importantly, the sandbox is complemented by a blacklist. Furthermore, like similar Android application sandbox approaches (e.g. [45, 124, 125]), the system can help in detecting potential misbehaviour by third-party apps. Judged by the sandbox automatically, apps with potential malicious intents are marked as 'suspect malicious apps' and reported to users for a final judgement.

With support from the sandbox mechanism, and manual input from users, installed apps can be added into the whitelist or the blacklist. Apps in the whitelist will be completely trusted and be able to obtain all information they require, while all permissions of apps in the blacklist will be withdrawn so that they can obtain only mocked information.

5.3.4 Simulation and evaluation

The test device for our prototype is a LG Nexus 4 (1.5 GHz quad-core Krait, 2 GB of LPDDR2 RAM clocked at 533MHz), which runs a rooted firmware based on Android 4.3.1. The software testing environment consists of 200 apps in total: 60 off-the-shelf apps selected randomly from the top free apps of the Google Play Store; 69 built-in apps that come with the firmware (some of the apps can be operated by users, for example, Gmail and Google Map; the others, such as Themes Provider and Calendar Storage, are hidden from users); 59 apps for daily use, including both free and paid apps; 10 dummy apps that consume ads from testing cooperative

Table 5.3: Cost of the prototype

Term	Average	Sample variance
Memory consumption	9066KB	4855471 (n=10)
Power consumption in standby mode for 24 hours	less than 1%	N/A
Ad scanning time per app	24.82 ms	799.23 (n=200)
Time of obtaining real location	5.19 ms	124.39 (n=100)
Time of obtaining and faking location	5.91 ms	114.97 (n=100)
Time of sending a SMS message without monitoring	4.05 ms	5.20 (n=100)
Time of sending a SMS message and recording details	4.81 ms	6.91 (n=100)
Time of loading original ad url from dummy ad-network	0.01 ms	0.009 (n=100)
Time of replacing original url and loading new local ad url	0.02 ms	0.019 (n=100)

ad-networks built to work with our system; one app as the PPTMA client that can be used to adjust related configurations; and one app as the client of Cydia Substrate that helps to accomplish API hooking. The details are shown in Table 5.2.

According to the Android memory management mechanism, if the system runs low on memory, processes may be killed in the LRU cache beginning with the processes least recently used. In order to keep a stable testing environment, we select only three apps from each of the first four types of apps to run, with one dummy app which consumes banner ads from the dummy ad-network running in the foreground. Each dummy app tries to get a banner ad from the dummy ad-network after it starts. It also tries to get the location from the device and send an SMS message to the ad-network without telling the user in every second. When the PPTMA system works, APIs related to obtaining location and sending SMS messages are hooked: fake locations are submitted to the apps; operations of sending SMS messages are monitored; and details such as the destination, the source, the text content, the date, and the process ID of the app are recorded.

We first evaluate the costs of the prototype in terms of the memory and power consumption, then observe its impact on the performance of the third-party apps, and, finally, we evaluate the efficiency of ad-scanning. We measured time costs of API hooking and ads replacing by averaging over several runs of one dummy app which contains job timer in its code. Test data is shown in Table 5.3.

As expected, the costs of memory and power consumption are relatively low compared to the totality of the available resources. In addition, there is a performance overhead of only 13.9% (from 5.19 milliseconds to 5.91 milliseconds) after the location API hooking on third-party apps, which is also acceptable. The overhead pertains to stopping apps from getting real data from GPS sensors and replacing the return value with fake data. A performance overhead of 32.9% occurs after monitoring and recording the operations of sending SMS messages. Although 32.9% seems relatively high, the base number is actually very small (4.05 milliseconds).

The time consumption of ad scanning is currently quite high: it costs about 5 seconds to scan 200 apps. However, since the full scanning only happens when users scan their apps for the first time, we would argue that the waiting time is tolerable. The storage sizes of sample apps range from 0.53MB to 144MB, which contributes to the high sample variance. In addition, an incidental result shows that at least one ad-network plugin was found in 46 out of the 60 apps selected from the top free apps of the Google Play Store.

We also utilise the dummy apps to test the function of replacing original ads obtained from ad-networks with new ads selected from the local ads pool inside the mobile device. The result suggests that the information of ads can be changed and recorded successfully by PPTMA once we get the method names of loading ads. The progress can become much easier if we can obtain the limited versions of the SDKs from cooperative ad-networks — so that we do not need to spend time on detecting, hooking, and modifying their ad APIs. To clarify, time costs of loading and replacing ads mentioned in Table 5.3 only pertain to the operation of modifying the url of targeted ads. Since the prototype is only a starting point for us to develop and refine local ad selection algorithms, evaluation of selecting the most relevant local ad for users is not an issue at this time.

The prototype hooks location-related and SMS-related APIs. The former validates the functions of permission and personal profile management, while the latter validates the sandbox mechanism. Testing on app compatibility of the API hooking mechanism was performed, with the test data shown in Table 5.4.

Table 5.4: Compatibility testing

Term	Total number of apps	Passing number of apps	Passing rate
Location-hooking	46	46	100%
SMS-hooking	15	13	86.7%

The location-hooking mechanism functions properly in all of the 46 apps that require permissions for obtaining location, including some popular apps such as Facebook, Instagram and Google Maps. In terms of the SMS hooking mechanism, 15 apps hold the permission of sending SMS messages. However, in two of them we could not find a related interface to perform the sending operation. Monitored by our system, the two apps did not send SMS messages in the background. Although it is likely that they over-claimed the permission of sending SMS messages, we assume that they can send SMS messages silently and avoid our monitor because it is not obvious how to trigger the SMS sending progress in these apps. The SMS hooking mechanism works well in the remaining, with a passing rate of 86.7%.

We must acknowledge that, with more APIs being hooked by our framework, as well as more cooperative ad-networks being imported to it, there might be a corresponding increase in the consumption of memory and power. In addition, the performance of hooking APIs and obfuscating information might also be affected. However, since these new features are designed as individual add-ons for our main framework, and the system resources are mainly consumed by the framework, the influence of new add-ons on consumption and performance is expected to be tolerable.

5.3.5 Privacy improvements introduced by PPTMA

We now briefly discuss the privacy improvements introduced by our solution. The simulation of the prototype clearly demonstrated that, compared with the 'take-it-or-leave-it' strategy provided by the original Android system, our proposed architecture provided a more flexible mechanism to help users to have control over their personal information. More specifically, users can preserve their personal information in

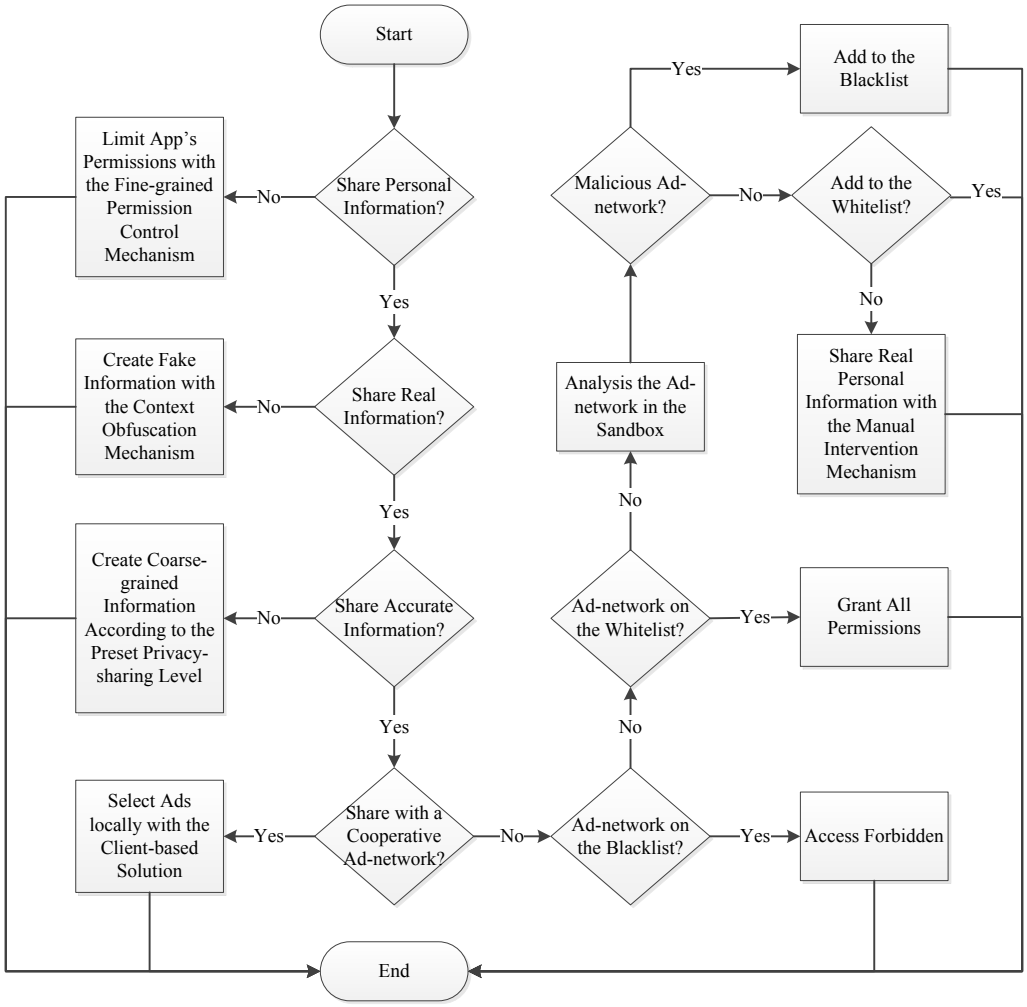

Figure 5.5: Personal information process flow

different levels and retain useful services of third-party apps with the prototype. A related flow chart of privacy context handling strategies is shown in Figure 5.5.

1. At a high level, a user can decide whether or not to share their personal information with particular ad-networks. If they choose not to share anything, they can withdraw all permissions relevant to their personal information from those apps. These functions are supported by the fine-grained permissions control mechanism in the permissions management module.

2. If the user chooses to share their personal information, no matter real or fake, to enable third-party apps to run in the way they are designed, they can decide whether or not to share real data. A context obfuscation mechanism

in the personal profile management module makes it possible for the user to create and submit fake information.

3. If the user chooses to share real information, then they can share real but coarse-grained information — for example, network-based approximate location instead of GPS and network-based precise location. This feature is supported by the permissions management module and the personal profile management module.

4. If the user chooses to share accurate information and the involved ad-network is one of those willing to gain consumers' trust and to work with our architecture, the sensitive personal information will be kept inside the mobile and associated with selected relevant ads locally.

5. If the given ad-network is not a cooperative partner, then it will be checked with the blacklist and the whitelist mechanism or monitored in a sandbox. For those apps not associated with these mechanisms, a manual intervention mechanism can be applied to them. For example, an explicit prompt will appear to ask the user whether the specific data could be shared with the ad-network when personal information is requested.

5.4 Summary

In this chapter, we have described the PPTMA architecture, which we have designed to be a consumer-targeted but business-palatable approach that helps consumers take advantage of mobile ads without compromising their privacy. We have described a prototype implementation based on the Android system that can be dynamically configured to hook sensitive APIs at run-time to enable mobile users to take control over their sensitive information.

5.4.1 Contribution over previous solutions

In terms of privacy protections, there are some similarities as well as differences among PPTMA and previous privacy-preserving advertising solutions such as Adnostic [3], Privad [24] and MobiAd [14]. We now briefly discuss them as below.

The core commonality among these solutions is described as the hybrid personalisation mechanism: a set of ads are downloaded to the client and the best ad is then selected with fine-grained profile stored locally. The main differences, on the other hand, lie among the design principles of each solution.

Adnostic and Privad are designed to work on desktop PCs instead of mobile platforms, while PPTMA works mainly on mobile devices. In addition, ad-networks' knowing of user clicked on an ad is not considered as a privacy violation in Adnostic, while PPTMA takes into account a click report as it could be used to infer users' preferences. Meanwhile, Privad is a purely technological approach which does not trust any ad-networks and anonymses every piece of user data. Compared with Privad, PPTMA combines technological and policy based approaches. It allows users to deploy the solution by themselves as well as support innovative uses with cooperative ad-networks.

Both MobiAd and PPTMA are designed to work in a mobile environment. However, the main focus of MobiAd is local ads — local mobile base stations and local WiFi hotspots are used to deliver ads. Compared with MobiAd, PPTMA performs ad delivery via remote servers and makes use of TRE to process click reports and detect click fraud.

Additionally, compared with the above-mentioned solutions, PPTMA is not only a privacy-preserving advertising system, but also a user-centric management system for users to interact with their own data.

5.4.2 Possible objections

A number of questions come to mind with respect to this solution.

1. **View/click report processing.** Ad-networks, advertisers and content providers need the click records of ads for charging money and sharing payment. Although we can keep users' personal data inside the mobile device by applying a local ad selection mechanism, we still need to download the selected ad and provide the click report, which, in turn, can be analysed by the ad-network to infer the user's interests. To provide the click records without exposing which user performs the particular operation, we propose the application of the Trustworthy Remote Entity (TRE) [126] architecture. The TRE is a computational and communication system that performs online processing of information provided by two or more communicating parties. In our proposed framework, the click records of individual users are first delivered to a TRE; the TRE then hides user IDs from different records and delivers the result to ad-networks with single-use random IDs.

2. **Click-fraud.** In current TMA systems, ad-networks need the ad click reports of consumers to defend against click-fraud. In our model, only single-use IDs that are randomly generated by TREs will be submitted to ad-networks with click reports, which limits the ability of ad-networks to detect click-fraud. However, the mappings between the submitted IDs and real user IDs are stored in the TRE. Thus, click-fraud detecting can also be performed by TREs. TREs can help to analyse suspicious records such as a single user clicking on the same ads a number of times in a row, or different users clicking on a single ad in the same period of time, while the ad-networks can analyse other suspicious aspects such as the abnormal conversion rates of particular ads.

3. **Motivating advertisers.** There are a number of ways to motivate advertisers. First, since the information is edited by users themselves, rather than collected passively by unknown third-parties, users are more likely to

submit authentic data. In addition, as users know that the information is specifically used for selecting useful ads and the personal information is safe inside their mobile devices, they are more likely to submit accurate data voluntarily. Second, many organisations and governments have published policies to limit the ability of advertisers to track users. However, with our framework, advertisers will not need to track users. Third, since no personal information will be leaked when applying our algorithm, users will reduce the hostility towards advertisers, which, in turn, could increase response rates and transactions.

4. **Motivating users.** Though our system aims to provide balance, a potential problem is users ignoring most of the functions, and submitting fake data and blocking ads. Even if ads are displayed normally, users may still not read them if they have little interest [127] or if they have serious privacy concerns [74]. Furthermore, privacy advocates also have the potential to put an end to advertising models [128]. On the other hand, Barutçu [129] suggests that users are more likely to have positive attitudes to mobile advertising if they are price-conscious or more involved. A study conducted for Nokia by HPI Research [130] further identified key factors contributing to the acceptance of mobile advertising, which includes independent choice and mutual benefit. Findings from other studies (e.g. [131]) also suggest that consumers would be more open to receiving ads if they can obtain recognisable benefit such as special offers or discounts, and a relatively high tolerance to targeted ads can be achieved if the information of ads is perceived as useful. Furthermore, the appropriate adoption of a compensation mechanism for sharing personal information can also encourage users to apply the system.

5. **Business audiences.** Large companies will not give up their own ad-networks and apply our proposed solution. However, individual advertisers and small ad-networks who hold only limited resources could make use of a solution such as this.

6. **Security issues caused by root permission.** Our framework requires rooted devices. It is possible that more information could be stolen by malicious apps because of the root permission. To address this, we provide a compensation mechanism — a sandbox to monitor the behaviour of third-party apps. The sandbox could detect and track the information leakage caused by malicious apps. Thus, users can take action in time. Compared with other methods, such as app decompiling or system modifying, our choice of rooting devices and hooking APIs is the most simple method for users to apply privacy-preserving tools. Nevertheless, we will keep exploring the possibility of applying our solution without root permission.

7. **Ad blocking.** The fact that mobile users could make use of our solution to block ads from particular ad-networks or apps would make advertisers unhappy. However, even in the settings for Google Ads, users can choose to opt-out of interest-based ads on Google and opt-out of interest-based ads across the web. Apple also provides an ad-blocking feature in iOS 9 for users to block ads on their mobile devices. To establish a healthy environment, we should give users the choice to opt out of the personalised experience if they would like to. This could, in turn, drive advertisers to improve the quality of their ads.

8. **Permissions dilemma.** If a weather app requires location details both for the local weather and for targeted ads, users may struggle with the relevant permissions. In most cases, location information cannot be used alone to recognise a user: third-party apps need to collect the user's information, together with an identifier such as the device ID. Otherwise they may need to assemble many different attributes, such as mobile module, system version, location and feature of installed apps to verify a user. Thus, we can offer suggestions such as 'never submit Device ID' or 'do not submit a particular combination of information'.

9. **User acceptance of profile management.** It is reasonable that user acceptance of such a profile management feature would be influenced by factors such as the difficulty of maintaining it, the scope of applying it, the costs of learning it, and so on. In terms of providing user-centric management, some features provided by PPTMA are similar to previous projects such as Microsoft CardSpace (codenamed InfoCard, failed in February 2011[1]). However, the key difference here is that CardSpace attempted to expand its application to diverse online services — from relatively low risk services such as blogs or portal websites to more sensitive scopes such as bank accounts or access to official government websites. Such a wide scope requires the involved service providers to modify certain system mechanisms, to rewrite user agreements or contracts, to massively change the way they providing services in order to adapt different kinds of services to a unique profile management platform. In addition, the blurred and numerous concepts involved in CardSpace made it difficult to be accepted by users without related IT knowledge. Compared with CardSpace, PPTMA focuses on similar TMA systems and leaves the original mechanism of ad-selection unchanged, which makes the system easier to be maintained. The UI and logic of PPTMA is similar to those of general apps on Android system, which means users are able to make use of related experience to interact with PPTMA.

With these questions in mind, our solution provides a trusted platform that offers users a centralised management approach for users' personal information and provides a fine-grained access control mechanism for controlling installed apps. The system also support innovative uses with cooperative ad-networks. However, our proposed system also has the same challenges as faced by a databox [120]: availability, data complexity, cost, and so on. In addition, we must acknowledge that we cannot protect all personal information, even if we were to combine all existing privacy-preserving solutions: we cannot prevent Apple's algorithms from knowing which apps users have downloaded and installed; we cannot stop Google's

[1] `https://blogs.msdn.microsoft.com/card/2011/02/15/beyond-windows-cardspace/`

algorithms from analysing emails in Gmail and calendar events in Google Calendar. It is unarguable that, in order to preserve consumers' privacy appropriately, technical and legal solutions, as well as education, also have a role to play.

Our prototype implementation provides us with a starting point to start to develop and evaluate mechanisms for selecting ads on the basis of user preferences and permissions. In the next chapter, we will discuss AdSelector as an instance of the configurable ad-selection mechanism of PPTMA.

6

AdSelector: A Privacy-Preserving Advertisement Selection Mechanism for Mobile Devices

Contents

6.1 Overview of AdSelector 128
6.1.1 Coordinate with PPTMA 128
6.1.2 Ad interest categorisation in AdSelector 130
6.2 System architecture 131
6.3 Core mechanisms and workflow 134
6.3.1 User subscription mechanism 137
6.3.2 Two-stage ad selection process 139
6.3.3 Privacy-conscious billing 144
6.3.4 Advertising fraud defence 146
6.4 Simulation and evaluation 150
6.4.1 Simulation setup 150
6.4.2 Simulation scenarios 151
6.4.3 Privacy improvements introduced by Ad-Selector 156
6.4.4 Evaluation of performance 156
6.4.5 Limitations and challenges 159
6.5 Summary 160

In the previous chapter, we described the design of the *Privacy-Preserving Targeted Mobile Advertising (PPTMA)* framework. A key feature of the framework is that it can be deployed by users, and is also intended to be palatable to businesses

that operate in this space. This feature is enabled by supporting a configurable ad-selection mechanism module that can be customised by advertisers. To this end, we now present the design of *AdSelector*, a profile-based advertisement selection mechanism for mobile advertising, as an instance of the configurable ad-selection mechanism module that is able to coordinate with PPTMA.

The contents of this chapter are based on *Yang Liu and A. C. Simpson: AdSelector: A privacy-preserving advertisement-selection mechanism for mobile devices. In: The Computer Journal 60.8 (2017), 1251 — 1270* [132].

6.1 Overview of AdSelector

The fundamental driver of AdSelector is the need to exchange coarse-grained information pertaining to ads from the ad-networks with very limited anonymous user data, and selecting the most useful ads with detailed user profiles on the mobile clients. To provide protection for both users' and ad-networks' benefits, several mechanisms are combined. We start by introducing the way that the core mechanisms coordinate with the PPTMA framework.

6.1.1 Coordinate with PPTMA

As discussed in Chapter 5, PPTMA serves privacy-friendly targeted ads in two ways: via either a 'cooperative mode' or a 'mandatory mode'. In the cooperative mode, ad-networks are aware of PPTMA, and tailor-made ads are selected with users' local information on the client. In the mandatory mode, the system hooks API calls that can expose users' personal information, and consults a local policy to decide how to proceed.

As a general advertisement selection mechanism, AdSelector can be deployed in PPTMA to implement the 'Configurable Ad Selection Mechanism' module in the cooperative mode. Figure 6.1 presents an overview of the architecture of PPTMA and indicates the position of AdSelector.

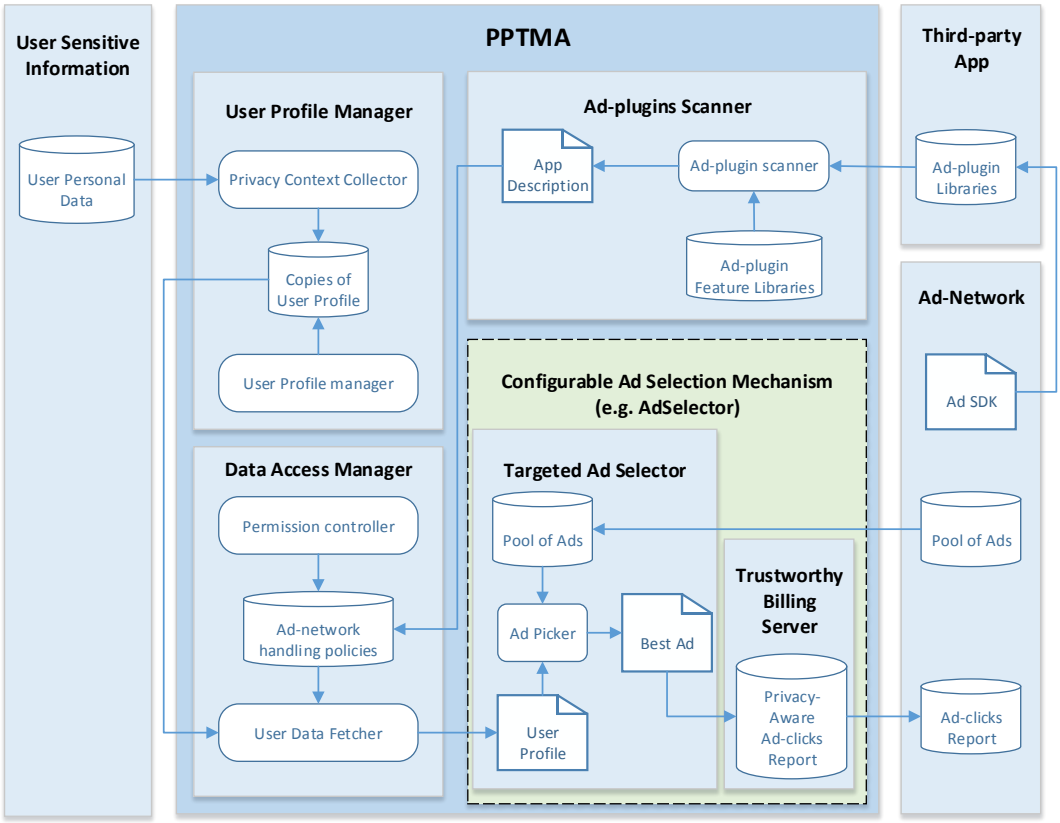

Figure 6.1: Architecture overview of PPTMA

The main feature of AdSelector is its combination of a user subscription mechanism, a two-stage ad selection process, and the application of a trustworthy billing system. In particular:

- The user subscription mechanism helps users to identify their interests and subscribe to desirable categories of ads.
- The two-stage ad selection process ensures that ad servers can obtain only coarse-grained user profiles, with fine-grained user profiles stored and used only on the mobile devices.
- The trustworthy billing system helps to report ad-views and ad-clicks without revealing users' identities and can assist in detecting advertising fraud attacks.

As shown in Figure 6.1, the user subscription mechanism coordinates with the user data fetcher of PPTMA to maintain user profiles; the two-stage ad selection

process works on the pool of ads on the server of an ad-network to create local ads; the trustworthy billing system sits between the local ad-clicks report dataset and the corresponding dataset of the ad-network to provide assistance.

6.1.2 Ad interest categorisation in AdSelector

We now briefly introduce the ad interest categorisation adopted in our solution. In most TMA systems, advertisers are required to specify particular ad interest categories when publishing ads. Users involved in these TMA systems are also associated with one or more of these ad interest categories — either deduced automatically by ad-networks or specified by users themselves — to indicate the kinds of ads and content they are interested in. The ad-networks then give consideration to some other factors (e.g. advertisers' budgets, users' ad preferences and location, etc.), and deliver particular ads to related users who are classified in the same categories.

In general, ad interest categorisations have a hierarchical structure. For example, at the time of writing (September 2017), the Google ad platform [119] uses a multi-level categorisation consisting of 25 top-level categories and up to 7 sub-levels. By way of comparison, the Yahoo ad platform [133] uses a categorisation consisting of 16 top-level categories and 5 sub-levels. Finally, the Facebook ad platform [134] uses a hierarchical ad interest categorisation consisting of only 9 top-levels. However, apart from the interest categorisation, Facebook also applies a hierarchical demographics categorisation (e.g. education background and income condition) and a hierarchical behaviours categorisation (e.g. purchase behaviour and trip frequency) to help profile users.

Despite the different classification criteria used in the above ad platforms, they all apply the same categorisation structure. The hierarchical structure makes a contribution to characterising users and ads in an effective way. A user or an ad can be assigned to different levels of the categorisation according to the particular requirement of accuracy or according to users' data access preferences. Additionally, the feature of assigning an item to more than one category provides flexibility.

As an example, we write A > B to represent the notion that B is a sub-level item of A. Thus, a typical category of the Google ad interest categorisation might be the three-level item Sports > Individual Sports > Golf. Any ads about golf, for example, pertaining to equipment, video games, or tickets can all be associated with this category. If the advertiser wants to narrow the scope to make their ads more precise, they can then specify an additional category Sports > Sporting Goods > Golf Equipment for ads for golf clubs and assign Games > Computer & Video Games > Sports Games to ads for golf video games. Meanwhile, according to a user's data access control policy, they can be associated with different levels of the item — Sports > Individual Sports > Golf, Sports > Individual Sports, or, even more coarsely, with the top-level category Sports.

6.2 System architecture

The fundamental mechanisms suggest that there are two key interactions between our system and ad-networks:

1. AdSelector provides coarse-grained user information to ad-networks to exchange potential lists of ads.

2. Ad-networks obtain trustworthy but privacy-conscious ad view/click reports from AdSelector to support billing.

At a high level, our system can be implemented in two distinct ways: as a third-party app or as an external library. In the first mode, users install the system on their mobile devices as an app. Ad-networks need to provide only simple APIs to interact with the system instead of modifying their existing facilities. In the second mode, users are not required to install any extra apps. However, ad-networks need to import an external library into their ad SDK, and obtain coarse-grained users' data and view/click reports by calling related APIs provided by the external library.

There are some trade-offs associated with both modes. In the third-party app mode, ad-networks do not need to modify the way they serve ads. Following the

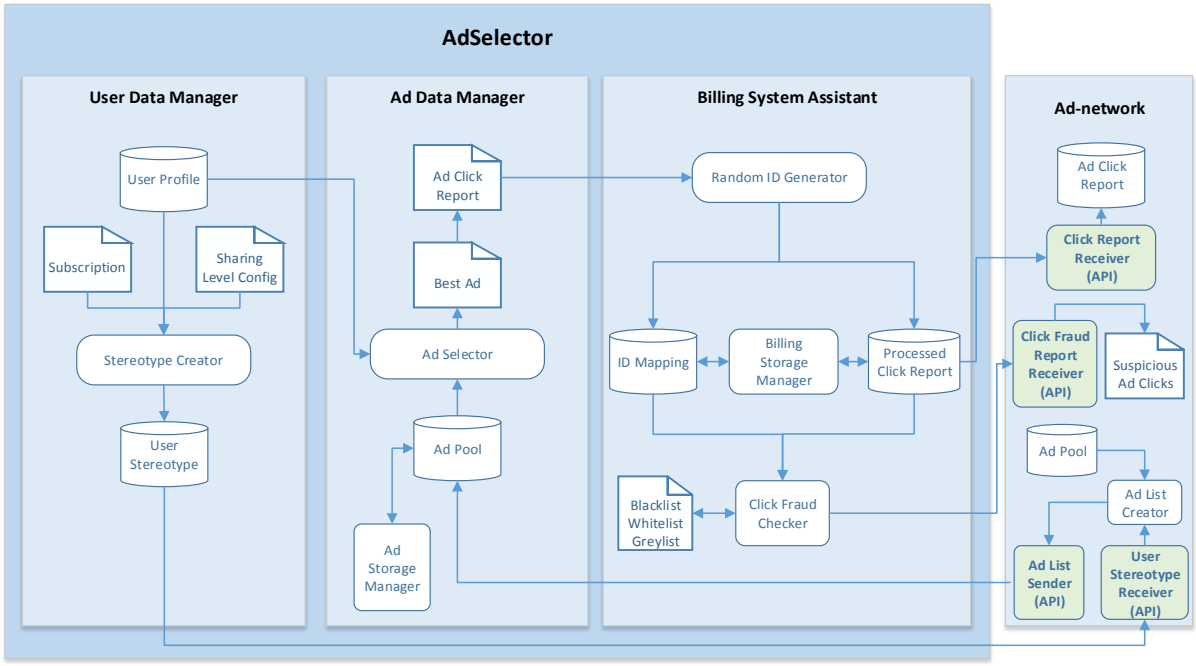

Figure 6.2: System architecture

installation, users' devices can detect related ad-networks automatically and call the light version APIs to accomplish the key interactions. If the app of AdSelector is not installed, ad-networks can serve ads via their existing processes. In the external library mode, ad-networks can serve ads in the privacy-friendly way to all mobile users. However, they would need to modify their existing codes and servers to apply the new features.

Figure 6.2 shows the system architecture of AdSelector for the first of these modes. There are three main components:

1. **User data manager**. This module maintains all kinds of user data (e.g. device ID, location information, installed apps, etc.) that are collected from the mobile device. User profiles are generated in terms of the collected data. In addition, together with the information of particular ad categories that users have subscribed to and the preset access control policies, different copies of user stereotypes — the coarse-grained anonymous user data that helps ad-networks to generate potential lists of ads — are also created by the user data manager.

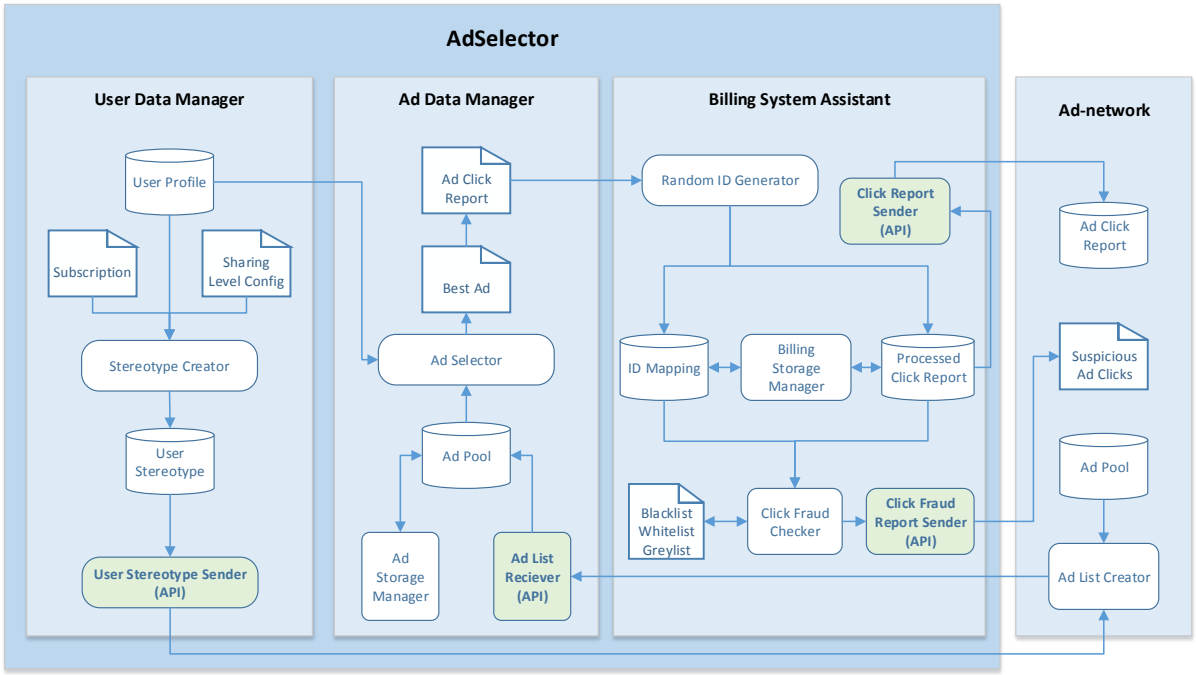

Figure 6.3: Alternative system architecture

2. **Ad data manager**. The ad data manager maintains a local pool of ads. The lists of ads in the pool are obtained from different ad-networks with different copies of user stereotypes. A storage manager helps to optimise the pool of ads — removing, for example, the ads that meet certain conditions, e.g. an ad that has been displayed many times but never clicked. The ad selector filters ads from the pool of ads and selects the most relevant ad for the user based on their detailed and contextual user profile. The original data associated with ad view/click reports is also stored in this module.

3. **Billing system assistant**. This module helps to process ad view/click reports before they are delivered to ad-networks. A single-use user ID is generated by the random ID generator to replace the real user ID inside a new view/click report that is received only by this module. The mapping of the two kinds of user ID are stored in the system for detecting view-fraud and click-fraud attacks, while the view/click reports with single-use user IDs are sent to the related ad-network for billing advertisers and paying app developers. In addition, a billing storage manager is provided to control the

retention and deletion periods of the mappings and ad view/click reports. There is also a blacklist–greylist–whitelist mechanism built into this module that is used to mark different levels of users' behaviours.

Third-party apps connected with ad-networks can interact with the system in either of the two modes mentioned above. Figure 6.2 represents the first mode: users install the app of AdSelector on their mobile devices and ad-networks provide related APIs to implement the ad list creator and the view/click report fetcher to perform the interactions. The second mode works similarly, but the related APIs are provided by the AdSelector library. The alternative system architecture is shown in Figure 6.3.

6.3 Core mechanisms and workflow

The core concept of AdSelector involves selecting the most relevant ad with detailed user data being stored on users' devices, rather than on ad-networks' servers. Some previous approaches, such as Adnostic [3], Privad [24] and MobiAd [14], have similar goals. The novel feature of our system is the combination of several mechanisms that help to improve ad selection effectiveness, to increase user participation, to provide billing and accounting function in a privacy-preserving fashion, and to defend against click-fraud attacks. We consider each in turn.

1. **User subscription mechanism**. Users are able to identify their interests and subscribe to specific categories of ads. The mechanism has the potential to increase users' tolerance to targeted ads and increase their engagement in targeted mobile advertising.

2. **Two-stage ad selection process**. The most relevant ads for a particular user are selected via a two-stage process. Users' subscription profiles and related coarse-grained information are first delivered to ad-networks anonymously for pre-downloading lists of ads that the user may be interested in. Then the fine-grained user profile stored in the mobile device is used locally to select the best ads to display. The two-stage selection process helps to keep users' private data inside the local device.

3. **Privacy-friendly billing system**. Ad-networks need the view or click records of ads for billing advertisers and paying app developers. The view/click records, in turn, can be analysed to infer users' interests. The privacy-friendly billing system we provide is an instance of the Trustworthy Remote Entity (TRE) [126], a highly-specialised networked system with very minimal function for data processing. It helps to identify which ads have been displayed by which apps without telling ad-networks which users performed the views or clicks.

4. **Advertising frauds defence mechanism**. Advertising frauds, particularly view-fraud and click-fraud, cause serious damage to the TMA ecosystem. Undetected click-fraud can cause large losses to advertisers — Pearce *et al.* [135] claim that advertising losses of about US $100,000 per day can be caused by a single click-fraud botnet (ZeroAccess). Further, Dave *et al.* [136] suggest that around one third of mobile ad-clicks are suspected to be click-spam. Hence, there is a need to detect mobile ad click-fraud in every ad-network system. The fact that ad-networks cannot identify users in our solution may reduce the ability to detect advertising frauds; we propose several approaches to resist such attacks as a compensation mechanism.

The system obtains the ability to perform privacy-preserving targeted mobile advertising by combining all of the above mechanisms in terms of the workflow of Figure 6.4. This workflow is described below.

1. A user subscribes to particular ad categories that he or she is interested in, and presets the data sharing levels of all installed apps (i.e. which parts of the user information can be obtained by which third-party app).

2. The information provided by the user, together with detailed user data that is collected by our system, is then used to generate user stereotypes in the form of coarse-grained anonymous user data.

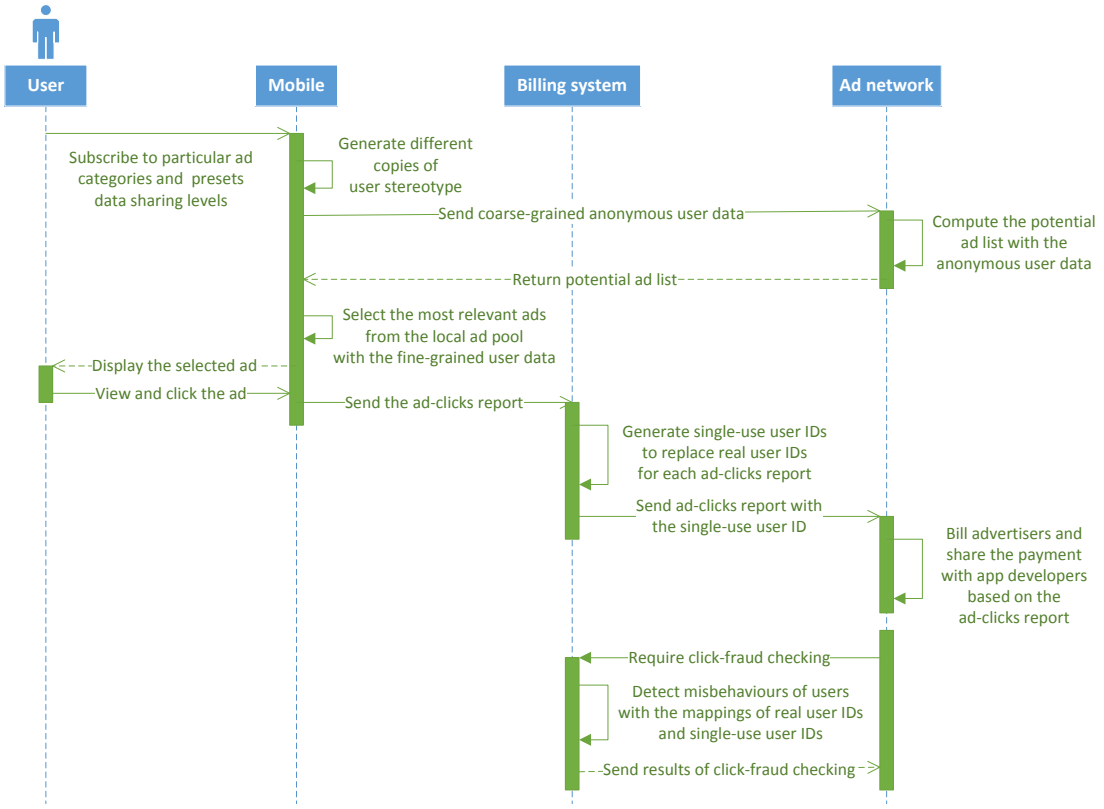

Figure 6.4: System core workflow

3. A particular copy of the user stereotype is sent to a specific ad-network. The copy is then used to select ads that the user might be interested in and generate the potential list of ads in the server of the ad-network. The list of ads is sent back to the mobile device and stored in the local pool of ads.

4. On the mobile client, the most relevant ads are selected from the local pool of ads by computing the fine-grained user data. The selected ads are then displayed on the mobile device and the user can then view and click on them.

5. After the view/click of an ad, the view/click report with the real user ID will be sent to the billing system from the mobile client. In the billing system, a random single-use user ID is generated to replace the real one. The processed view/click report is then submitted to the server of the related ad-network, while the mapping of the real user ID to the single-use user ID is stored in the billing system for a period determined by the ad-network.

6. The ad-network can bill advertisers and share payment with app developers based on the view/click reports. When there is a possibility of view-fraud or click-fraud attacks, the ad-network sends a request to the billing system to perform related detection.

7. The billing system detects misbehaviours of users with the ID mappings and view/click reports, then submits results of view-fraud or click-fraud checking to the related ad-network.

6.3.1 User subscription mechanism

User profiles generated by ad-networks might sometimes be questionable or untrustworthy because users might obfuscate their data for the sake of privacy-preservation [137]. Additionally, the algorithms used for deducing users' interests might not be sufficiently effective — users understand their own interests and needs better than ad-networks working with incomplete users' profiles [138].

In the privacy-friendly TMA environment supported by AdSelector, users can take advantage of mobile ads without worrying about their personal information being misused. Furthermore, users can actively make efforts to optimise the targeting process to increase the accuracy of selecting relevant ads for their own benefit. Findings from previous studies [130, 131] suggest that independent choice and mutual benefit can contribute to users' acceptance of mobile advertising and increase their engagement in the targeting process. Hence, a user subscription mechanism is provided to give users the required abilities.

The following attributes can be customised by users. These attributes are chosen because they are generally collected and used by the likes of Google and Facebook.

1. **Age & Gender**. Personal information like age and gender is collected by most advertisers as these attributes may affect consumer response to advertising appeals significantly. With the subscription mechanism, a user can indicate her age with a range rather than with an exact figure. In addition, the mechanism enables users to decline to state their gender.

2. **Interests.** The interests of a user will be deduced by the user data manager of AdSelector and automatically added to the user's profile as they interact with the mobile device. In addition, users can edit the deduced interests and manually add keywords as specific interests.

3. **Location.** Users can decide how accurately their location data should be shared with different ad-networks. Additionally, they can subscribe to ads related to a specific location. For example, users planing to travel to London can choose to receive ads and discount coupons for there.

4. **Types of ads.** There are different types of mobile ads such as introduction of new products, discount coupons, public service ads, etc. Users can indicate which particular types they are interested in.

5. **Categories.** Generally, ads are published in ad-networks and associated with one or more categories. Accordingly, users can subscribe to categories of interest.

We have chosen to leverage Google ad interest categories. As one of the largest ad-networks serving ads on both Android and iOS platforms, Google makes its ad interest categories openly available. Applying Google ad interest categories has the potential to make it easier to adapt our mechanism to current TMA systems.

Additionally, users can set weights for the attributes to which they have subscribed. The data of users' subscriptions is maintained by the user data manager of AdSelector. According to the requirements of advertisement selection algorithms applied by different ad-networks, the subscription data can be presented in different formats. For example, Food & Drink > Food > Meat & Seafood > Poultry can be converted to the set { Food, Drink, Meat, Seafood, Poultry } for fuzzy matching — which ignores the hierarchical information so that all factors in the set are considered with the same weight. Alternatively, the category can be converted to the sequence ⟨ Food & Drink, Food, Meat & Seafood, Poultry ⟩ for more sophisticated matching. Here, the hierarchical order of factors in the sequence is considered. Thus, the

factors could be computed with different weights. Furthermore, if a related ad-network uses a fixed category list, the subscription information can be converted to a vector-based representation that enables simple comparison.

6.3.2 Two-stage ad selection process

Having generated the user profiles and collected the subscription information, further work can be done to complete the process of ad selection and display. To this end, we now provide details of the ad selection process applied in AdSelector.

The main idea of the ad selection process is to keep and use the fine-grained user profiles only on the mobile devices so as to reduce the possibility of users' information leaking. To accomplish this, we perform the ad selection process in two stages. The first stage decides which potential ads should be cached in the mobile clients; the second stage decides which particular ad should be displayed.

In the first stage, the coarse-grained version of a user's profile is delivered to the server anonymously. Based on the coarse-grained information, the ad-network can select a number of ads that the user may be interested in. These ads, together with some essential attributes, such as targeting location, gender and ad interest category, are then stored in an ordered list of ads. The list of ads is downloaded to the user's mobile client at the end of this stage.

In the second stage, the fine-grained version of the user's profile and the pre-downloaded ads are encoded in a common format to perform the comparison. All of the ads are scored and ranked according to how accurately they match the user's profile and the bid budget set on the ads. The most relevant ad at the top of the list will be displayed.

The default ad selection approaches for the two stages are also presented below so as to describe the complete workflow of AdSelector. These approaches can be replaced with pre-existing ad selection algorithms in both stages.

Pre-download: get the potential list of ads from an ad-network

The best way to perform the pre-download process is to download all ads from all servers of the ad-networks to the user's mobile device. Thus, the user would not lose any potential ads and the ad-networks would not learn anything about the user.

Although the ideal approach cannot be implemented simply due to many hurdles (e.g. storage, bandwidth, and performance), we strive to achieve an appropriate approximation. A suitable pre-download approach should meet the following requirements.

1. Users should be able to download enough relevant ads to be selected and displayed in the second stage.

2. Ad-networks should not be able to identify the users nor learn their accurate interests.

3. The approach should be compatible with approaches adopted by, for example, Google, Facebook, and Twitter.

4. The representation used in the approach should enable simple comparison in both stages.

A keyword-matching approach is chosen to be the basic ad selection approach in the pre-download process. On the one hand, most of the existing ad-networks support keyword matching, as it is one of the main approaches to filtering ads. On the other hand, user profiles generated in AdSelector can be converted easily to a set of keywords. Figure 6.5 shows an overview of the two-stage ad selection process based on the keyword-matching approach.

The information used to generate the user profiles are collected automatically and/or edited by the users manually via the subscription mechanism. Configured by the user, a copy of the user profile can be encoded into different keyword sets with different accuracies.

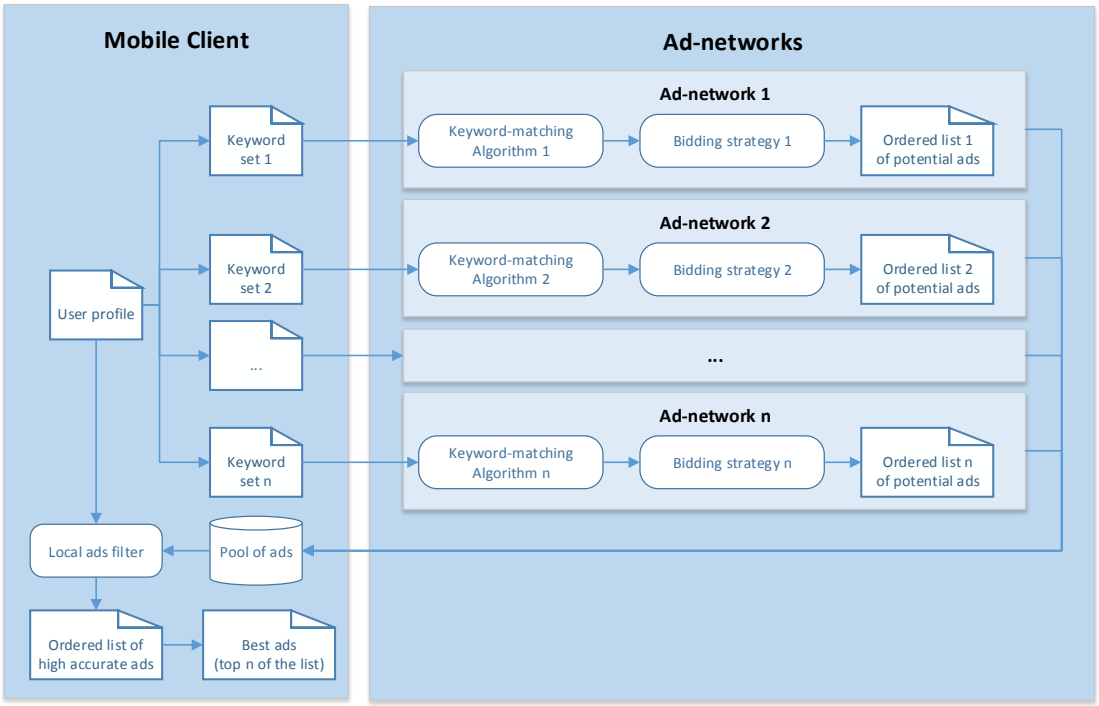

Figure 6.5: The two-stage ad selection process

An example copy of a user profile might contain the following information:

Gender: Male.

Age: 38 years old.

Interests: Football and reading.[1]

Location: Paddington, London, UK.

Types of ads: Discount coupons.

Categories: Food & Drink > Food > Meat & Seafood > Poultry.

Depending on the access control policies applied by the user, the keyword set could take one of several forms, such as { Male, 38 years old, Football, Reading, Paddington, Discount Coupons, Poultry } and { 20 to 40 years old, Team Sports, Books & Literature, London, Food }.

[1]As interests can be edited manually by adding or removing keywords via the subscription mechanism, any words can be associated with this attribute.

Note that in the second set, which is relatively coarse-grained, the user's interests are convert into keywords that are extracted from different levels of the ad interest categories we use: Football is extracted from Sports > Team Sports and Reading is extracted from Books & Literature; Poultry, the category manually subscribed to, is promoted to the higher level category Food to reduce the accuracy of the information to be submitted. In addition, the age and location are generalised, and other information, such as gender and the types of ads, are suppressed.

After receiving the keyword sets, ad-networks can execute their own algorithms to select potential ads. Although the keyword-matching mechanism is supported well in most existing ad-networks, many algorithms pertaining to keyword-matching or ad auctions are commercially sensitive and cannot be shared. In addition, it is also unrealistic for an approach such as ours to provide a universal ad ranking algorithm for all ad-networks.

Taking Google AdWords [139] as an example, some bids are set at the ad group level, rather than at the keyword level. Each ad group is associated with a set of keywords; for example, the keywords Solid-State Laser, Solid State Lasers and Gas Laser are all associated with the same ad group — Laser Systems. If a keyword from the set triggers an ad to appear, the price set on the ad group will be charged by the ad-network. Another example is negative keywords, which is also supported by AdWords. An advertiser who runs a pet food store but does not sell cat food can add *cat* as a negative keyword, so that the keywords *pet cat food* will not trigger the advertiser's ads to appear.

Since it is impossible to provide a unified algorithm that is compatible with the complicated bidding strategies and keyword-matching mechanisms behind all ad-networks, the approach only converts user profiles into keyword sets at the user end. The keyword-matching process can then be accomplished by individual ad-networks. Nevertheless, when generating the potential list of ads, the ads are required to be grouped by their trigger keywords with their original ranks maintained. Thus, in the local ads filter stage — the second stage of the ad selection process — AdSelector can

decide which ads are displayed in what order without knowing the particular bidding strategy of the ad-network and the bidding budget of the involved advertisers.

Other essential attributes (e.g. target attributes, ad type and ad format) of each ad are also stored in the list of ads for filtering out no matching ads in the second stage. For instance, the target attributes such as targeting location, gender and language help to filter out ads that do not match the fine-grained user profiles. The ad type and format help to select the most suitable ad view for the current screen to display (e.g. banner image ad and skippable video ads).

To reduce the consumption of storage and bandwidth in the pre-downloading process, users can set the number of ads to be downloaded each time and set the total number of ads to be cached in the mobile devices. In addition, only the essential attributes are downloaded to the list and stored in the cache of mobile devices. The long text descriptions and images of ads, which are expensive in terms of storage and bandwidth, are only retrieved from the ad servers when the relevant ad is selected and needed to be shown on the mobile device.

Local ads filter: select and display the most relevant ad on a mobile client

The process of the local ads filter stage is similar to the existing ad selection process utilised by ad-networks: the system holds a set of ads and tries to select the most relevant one by analysing the fine-grained user profile.

Realistically, ad-networks would not be prepared to disclose their ad selection strategies. As such, we describe a simple approach to show how the complete process of ad selection can still be accomplished.

A pool of ads is maintained on the mobile client to store a potential list of ads from different ad-networks. After an ad request is generated, AdSelector first picks up the list of ads of the involved ad-network from the pool. As the list is ordered and each ad preserves its own rank computed by the ad-network in the pre-download stage, AdSelector can simply filter out ads that do not match the fine-grained user profile, do not match the ad type and format of the ad view in the current page, and do not match the right cache duration. Thus, the ranking

information is still maintained in the shorter but more precise list. The client can then show ads from the shortlist according to their ranks.

In addition, some other mechanisms can be applied at the same time to revise the result. For instance, a time-weighted mechanism generates a different result from the same keyword according to the time of day (e.g. the result of searching for the keyword Food in the early morning may be different from searching for the same keyword at midday). The weight of different factors that a user sets via the subscription mechanism can also alter the ad selection result. (We do not discuss this further here, as our focus is the privacy-preserving mobile advertisement selection mechanism, rather than a specific ad selection algorithm.)

Apart from the local ads filter process, some strategies are also performed to maintain the pool of ads. For example, the ads view history is recorded to facilitate the removal of particular ads that are frequently displayed but rarely clicked.

6.3.3 Privacy-conscious billing

The key to enhancing the privacy of the billing system is to ensure that ad-networks can obtain the information of which ads are viewed or clicked in which apps, without identifying which users performed the operations. To achieve this, AdSelector utilises the Trustworthy Remote Entity (TRE) approach of [126].

A TRE is a computational and communication system that enhances privacy in communication exchanges. The TRE system is situated between two or more communicating parties to perform information processing. The TRE provides strong guarantees of its trustworthiness by using technologies and approaches from the field of Trusted Computing — the communicating parties verify the state of the intermediary to confirm its trustworthiness, rather than relying on a trusted third-party.

In our system, the operation data of ads is first processed by the TRE system. A privacy-aware version of the data is then submitted to related ad-networks.

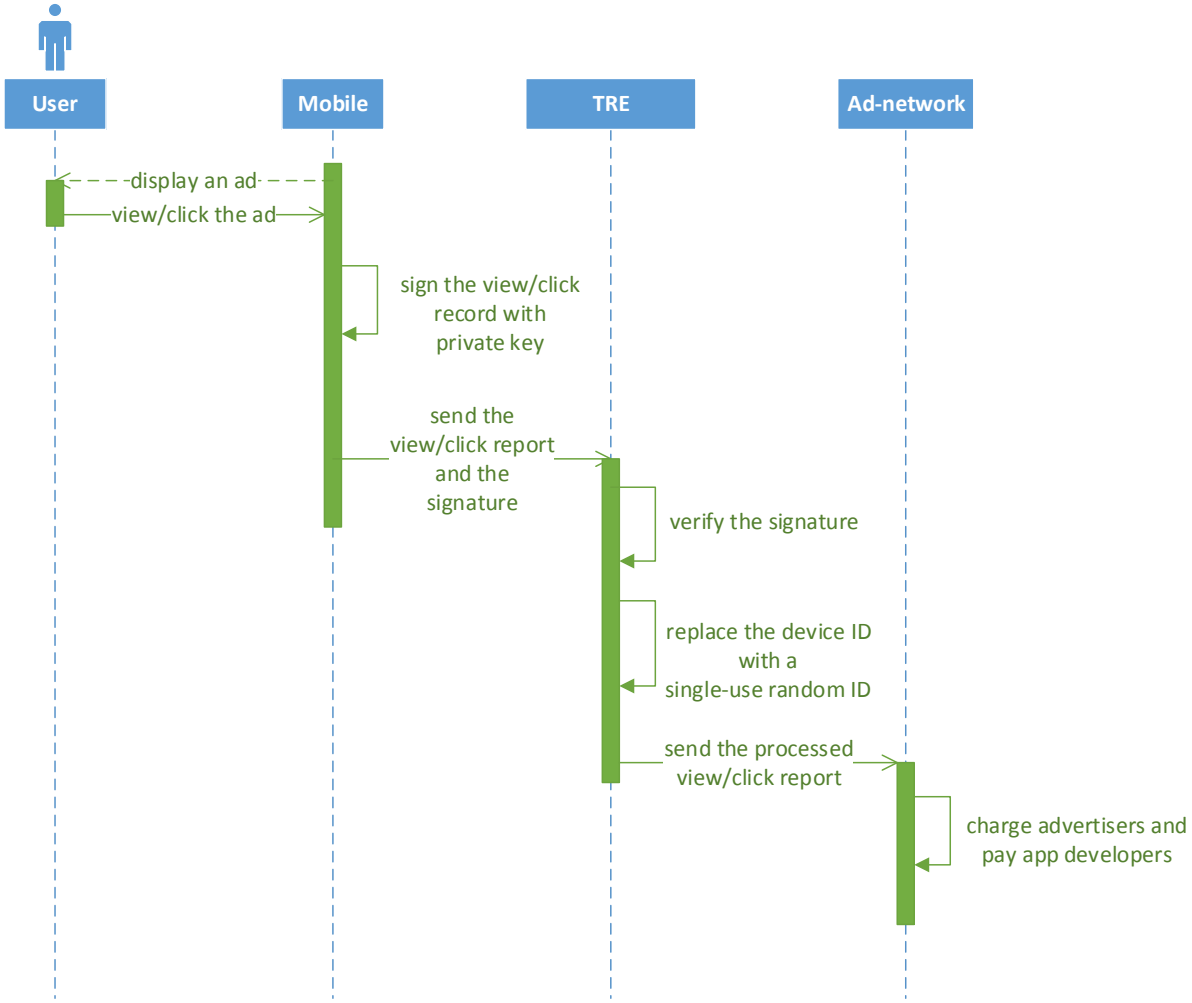

Figure 6.6: Flow of steps in processing the view/click reports

The billing process is briefly described as follows. A related workflow is illustrated in Figure 6.6.

1. A user views or clicks on an advertisement.

2. The view/click report with the real user ID is signed with the user's private key and then submitted to a TRE.

3. The TRE verifies the signature, generates a single-use random ID for the user, and stores the mapping of the real user ID and the generated single-use ID for the purpose of detecting advertising frauds or tracing malicious behaviours.

4. The TRE submits the view/click report with the generated single-use ID to the ad-network.

5. The ad-network uses the report as proof to bill advertisers and pay app developers.

6. The ad-network can verify the state of the TRE to confirm its trustworthiness in each stage of the process.

6.3.4 Advertising fraud defence

By importing the TRE system, AdSelector obtains the ability to record ad-views and ad-clicks without disclosing information about users. However, the existence of advertising fraud prevents the stakeholders of the TMA ecosystem from simply using these records as billing proofs.

It has been shown that advertising fraud, particularly view-fraud and click-fraud, are causing serious damage to the TMA ecosystem [140–142]. The main revenue models of TMA are threatened by advertising fraud in the following ways.

1. **Cost-Per-View** (CPV), also known as Cost-Per-Impression (CPI), involves an ad-network charging an advertiser when an advertisement occurs on a user's mobile screen, no matter if the advertisement is clicked or not. In general, this model is implemented on a Cost-Per-Mille/Thousand (CPM/CPT) basis, with the fee being charged for every 1,000 impressions of an advertisement. Frauds against the CPV model include: resizing an ad view to one pixel; covering an ad view with other content; positioning an ad view to a nonexistent coordinate; and listing dozens of ad views in the same page (some mobile ad-networks require that only one ad view can be shown on a single screen [143]). Thus, ad-impressions are recorded without the ads being shown to the users. Alternatively, users might be dazzled by countless ads shown on a single page.

2. **Cost-Per-Click** (CPC) involves the advertiser paying an ad-network when an end user clicks on an advertisement. Click-fraud is the main threat to the

CPC model, whereby an attacker can simply click on an advertisement by hand or via automated scripts to consume the advertising budget of related advertisers.

While other revenue models exist — including Cost-Per-Action (CPA), Cost-Per-Sale (CPS), Cost-Per-Lead (CPL), Cost-Per-Install (CPI) and Cost-Per-Hour (CPH) — our focus in this study is threats to the CPV and CPC models.

Defending against view-fraud

In the CPV model, an ad-network determines whether an impression of ads is reliable or not by considering several different factors. We abstract the related factors to form the content of an ad-view record with the following attributes.

1. **Ad ID and ad-network identifiers**. In the two-stage ad selection process of AdSelector, an advertisement is pre-downloaded before being displayed. The two identifiers of the displayed advertisement are used by the TRE to locate the source of the original copy and to identify the ad-network that needs to receive the related record.

2. **Ad type**. Ads are categorised into three types: raw text, image and video.

3. **Ad format**. Each advertisement can be associated with a particular format such as banner ads, interstitial ads, skippable video ads, and so on. Each format has its own valid range of size, location, and duration. For instance, a small banner of the ad-network AdSense [144] should be sized to 320x50 and always be vertically aligned at the top of the screen. An impression of a skippable video advertisement is only valid if its duration exceeds 30 seconds.

4. **Timestamps of ad pre-download, ad request, and ad display**. The three timestamps are used to attest the valid date and the corresponding information flow of the advertisement. They can be used to detect invalid displays and replay attacks.

5. **Duration**. The final duration of an advertisement is recorded when the ad view ends. In addition, AdSelector checks and updates the duration of a displayed advertisement every 3 seconds in case of abnormal shutdown of the app.

6. **Screen size and resolution**. This is used to compute valid size and position of the ads displayed on the particular mobile screen.

7. **Ad size and position**. This information is verified according to the ad format of the displayed advertisement and the configuration of the mobile screen.

8. **Device identifier**. The device identifier is the real user identifier in AdSelector. The TRE stores the mapping of the device identifier to the generated single-use identifier for a short time to detect malicious behaviour.

9. **App identifier**.

As illustrated in Figure 6.6, to authenticate the ad-view record, AdSelector signs it with the private key of the user on the mobile client and verifies the signature on the TRE server. If the credibility of the ad-view record is confirmed, the TRE will then replace the device identifier with a single-use random identifier and submit the verified record to the related ad-network. Thus, the ad-network obtains reliable ad-views for charging advertisers.

Defending against click-fraud

The record of ad-clicks in the CPC model is generated in a similar way to the generation of the record of ad-views in the CPV model. The following attributes of the ad-clicks record differ.

1. The **ad click timestamp** is added to complete the cycle of processing ads.

2. The **duration** in the ad-clicks record is the difference between the ad click timestamp and the ad display timestamp.

3. The **ad click location** records the position where the user touches the screen. The location of the click should be in the area of the displayed ad view to be valid.

In addition, ad-networks can predefine some policies to exclude some real but invalid click operations. For example, a user clicking on an ad view only half a second after it is displayed may indicate a lack of any genuine interest in the advertisement.

Additional defence mechanisms

The above defence mechanisms help to ensure that the ad-views and ad-clicks are performed by real mobile users instead of being generated by scripts. However, there is still a possibility that the clicks are performed manually by attackers. Hence, we provide some additional approaches such as pattern matching, blacklisting, and bait ads as compensation mechanisms.

The TRE stores the mapping of real user identifiers and single-use random identifiers for a short time. Therefore, the TRE system has the ability of backtracking users' operation sequences. If the ad-networks predefine some suspicious patterns of actions, the TRE can then perform pattern matching to detect certain types of frauds.

For instance, a single user clicking on several ads in a row or different users clicking on the same advertisement (or different ads, but published by the same advertiser) in a short period of time are both suspicious patterns for some ad-networks. Hence, the records of ad-clicks in these patterns might be ignored immediately by the TRE before submitting them to the ad-networks. If a user is confirmed or identified as a malicious user, the TRE system can add the real user ID to a blacklist and simply block all future ad-clicks from this user. In addition, ad-networks can make use of bait ads [24] or bluff ads [28], which are both targeted ads with irrelevant ad content, to detect suspicious users that have the potential to be added to the blacklist.

6.4 Simulation and evaluation

An initial prototype of AdSelector has been implemented in the context of the aforementioned PPTMA. The evaluation has been primarily concerned with establishing the feasibility of the approach, rather than, for example, concerning itself with the specifics of ad selection algorithms. The performance overheads of some important operations (e.g. local ad selection and click report obfuscating) have been given due consideration.

6.4.1 Simulation setup

The prototype consists of three modules: an ad-network server, a server for a billing system, and a mobile device.

1. **Ad-network server.**

   The main functions of the ad server are maintaining data pertaining to ads, selecting targeted ads, and reviewing ad-click reports.

   In a fashion similar to the approach taken by Guha *et al.*, who use a trace of Bing search ads to evaluate Privad [24], we manually chose 100 real ads from Google AdWords to comprise the test data. To enlarge the sample, we also generated 9900 dummy ads with random targeting attributes.

   The ad selection algorithm that runs on the server is based on five targeting attributes: gender, location, age, interests, and subscribed ad categories. Other factors, such as the remaining ad budget, the ad-publish date, and the bidding strategy are not taken into account.

   The ad server runs on a virtual machine, the configuration of which is: 2.2 GHz single-core processor, 2GB DDR3 RAM, Ubuntu 14.04.4 LTS.

2. **Billing system server.**

   The billing system is a TRE instance with very minimal functionality: it replaces the user ID of the click report with a random single-use ID and stores

the mapping of the two IDs. The billing system also runs on a virtual machine, with the same configuration as that of the ad-network server.

3. **Mobile device.**

   As mentioned in Section 6.2, AdSelector can be implemented in two distinct ways: as a third-party app or as an external library. In the initial prototype, the system is implemented as a third-party app.

   An app named PPTMA is installed on the mobile device; the app allows the user to manually edit the user profile and subscribe to interested ad categories. In addition, the user can decide whether to activate the privacy-preserving mechanism or not.

   A second app named App-with-ads is responsible for collecting user profiles and displaying ads, as per common apps with ad-plugins. App-with-ads interacts with the ad server and PPTMA by importing related SDKs provided by the systems.

   The test mobile device is a Moto Nexus 6, the configuration of which is: 2.7 GHz quad-core Snapdragon 805 processor, 3GB LPDDR3 RAM, Android 6.0.1.

6.4.2 Simulation scenarios

We now present three illustrative scenarios.

Scenario 1

The aim of our first scenario is to observe which ad would be selected for the testing user profile without AdSelector, how much personal information would be collected by the ad-network, and how much time would be taken by these processes.

In this scenario: the device ID of the mobile device is collected as the user's ID by calling the related Android API; the user's gender and age are collected by using a questionnaire built into App-with-ads; the user's interests are deduced by

analysing apps installed on the mobile device; and the user's location is obtained by calling the GPS sensor.

Users' information pertaining to gender, age and interests can be collected by applying different strategies, all of which have different time consumption. However, information pertaining to users' locations are collected by different ad-networks in the same way with relatively consistent time consumption. Therefore, we record the time cost in obtaining a location from a GPS sensor to represent the time cost of collecting a user's profile in this scenario.

The simulation was performed with 10 different user profiles. For the sake of brevity, we use only one of these profiles to illustrate the simulation:

ID: ZX1G52533F.

Gender: Male.

Location: Oxford, UK.

Age: 25.

Interests: Basketball and video games.

The following operations are performed in this scenario.

1. App-with-ads collects the testing user's fine-grained information and submits it to the ad server.

2. The ad server selects the best ad for the user and shows the ad on the mobile device.

3. The user clicks on the displayed ad and sends the ad-click report to ad server.

The results of the simulation are shown in Tables 6.1 and 6.2.

Table 6.1: Time cost in Scenario 1

Term	Average time	Sample variance
Obtain location from GPS sensor	4.18 ms	2.46 (n=100)
Select the best ad in the ad-server	949 ms	284.83 (n=100)
Submit original click report to ad-server	965 ms	290.49 (n=100)

Table 6.2: Selected ad and disclosed personal information in Scenario 1

Term	Result
Final selected ad	Basketball shoe store in Oxford
User's information in the ad server	All information of the fine-grained profile

Scenario 2

In this scenario the privacy-preserving mode of PPTMA is turned on, meaning that the user can manually edit their fine-grained profile and generate coarse-grained copies. The aim of this scenario is to check, with the intervention of AdSelector, if the final selected ads are consistent with the displayed ad in Scenario 1. We also observe the differences of the disclosed personal information and the time consumption between this scenario and Scenario 1.

The user's fine-grained profile in this scenario is that of Scenario 1. The coarse-grained copy of the profile is edited as follows.

Gender: Decline to state.

Location: UK.

Age: 21-41.

Interests: Team sports and games.

Subscribed ad categories: Games > Computer & Video Games > Sports Games.

Number of ads that are downloaded to mobile per ad selection: 50.

Table 6.3: Time cost in Scenario 2

Term	Average time	Sample variance
Obtain location from fine-grained profile	0.31 ms	0.24 (n=100)
Obtain location from coarse-grained profile	0.30 ms	0.29 (n=100)
Select and download list of ads in the ad-server	891 ms	247.03 (n=100)
Select the best ads in mobile	1.30 ms	1.58 (n=100)
Submit original click report to privacy-friendly billing system	867 ms	252.41 (n=100)
Obfuscate click report in billing system and send it to ad-server	931 ms	234.97 (n=100)

In Scenario 2, App-with-ads collects both the user's fine-grained profile and the coarse-grained copy by calling the PPTMA API, as the information is already stored in the PPTMA app.

The operation flow of this scenario is as follows.

1. App-with-ads collects the user's coarse-grained information and submits it to the ad server.

2. The ad server selects the potential list of ads in the server, and sends the list to the mobile device.

3. Based on the fine-grained profile, App-with-ads selects the best three ads from the downloaded list.

4. The user clicks on the displayed ad (one of the best three) and the ad-click report is sent to the billing system.

5. The billing system replaces the user's ID from the ad-click report with a random ID and sends the obfuscated ad-click report to the ad server.

The results of the simulation are shown in Tables 6.3 and 6.4.

Table 6.4: Selected ads and disclosed personal information in Scenario 2

Term	Result
Final selected ad 1	NBA2K16 video game on game.co.uk
Final selected ad 2	Basketball shoe store in Oxford
Final selected ad 3	Basketball hoop at Amazon-UK
User's information in the ad server	Obfuscated ID and the coarse-grained copy of gender, location, age, and interests
User's information in the billing system	Original user ID

Scenario 3

The aim of the third scenario is to test the click-fraud defence mechanism. The following operations are simulated:

1. The user clicks the ad without reading the content (the user clicks in 0.5 second after the ad is displayed).

2. The user clicks a bait ad.

3. Several users click ads from the same advertiser in a short period of time.

4. The user clicks one ad several times in a short period of time.

The ad server maintains detailed information of ads and advertisers, while the billing system server holds the real user ID associated with the ad-click reports. Therefore, the detection of click-fraud attacks requires the cooperation of both servers.

In particular, the obfuscated user IDs involved in the first three operations can be detected by the ad server itself with simple SQL statements. The ad server can then send the IDs to the billing system and requires the billing system to add the original user IDs to a blacklist. In terms of the last operation, the process of the detection is reversed: the involved users are first detected by the billing system.

6.4.3 Privacy improvements introduced by Ad-Selector

The simulations show that the ads selected by AdSelector (Scenario 2) are consistent with the ad selected in the original TMA system (Scenario 1). In addition, the final selected ad 1 in Scenario 2 suggests that a more accurate ad can be delivered to the user due to the user's subscription mechanism — the ad pertaining to a basketball video game corresponds to the subscribed ad category Games > Computer & Video Games > Sports Games.

The differences in the user's disclosed personal information between Scenario 1 and Scenario 2 suggest that only the coarse-grained user profile can be obtained by the ad server. Furthermore, the ad server cannot identify the real user involved in the coarse-grained profile because the user ID is obfuscated by the billing system.

The simulation result of Scenario 3 shows that the ad-click report obfuscating feature of the billing system does not affect its ability to detect click-fraud. Cooperating with the ad server, the billing system can support a variety of click-fraud defence mechanisms.

6.4.4 Evaluation of performance

The time cost of the operations involved in the first two scenarios suggests that, although the number of operations in Scenario 2 is twice that for Scenario 1, the performance overhead of the prototype is not significant. The time cost of obtaining a user profile with AdSelector is only 15% of doing that by calling the Android API and checking the GPS sensor. In addition, local ad selection is about 700 times faster than remote ad selection. A significant overhead is caused by ad-click report obfuscating. AdSelector divides the original ad-click report submitting process into two steps: submitting the report from the mobile device to the billing system, then submitting it again from the billing system to the ad server. Therefore, the cost of processing an ad-click report with AdSelector is about twice of that in the original TMA system.

Another issue to be discussed is the consumption of storage and network bandwidth. It appears that pre-downloading and caching a list of ads would

Table 6.5: Consumption of storage and bandwidth of different ad formats

Term	Size
Targeting attributes information & URL for one ad in the pre-downloaded list of ads	0.16KB
Long text description for raw text ad	0.7KB
Image for standard banner ad	28KB
Image for full screen ad	74KB
Max size image allowed by Google AdWords	150KB
10 seconds video for video ad	191KB
30 seconds video for video ad	566KB

cause additional consumption in our model. Compared to Adnostic [3], in which approach a list of n ads are fully downloaded, AdSelector only downloads the essential targeting attributes and the URL of each ad in the list. The rest of the ad data such as long text description, image or video, which is expensive in terms of storage and bandwidth, will be downloaded later when the relevant ad is finally selected for the user. Thus, the consumption of storage and network bandwidth, as well as the time cost, are reduced.

There are many popular formats of mobile ads, including raw text ads in minimalistic style, banner ads built with relatively small images, interstitial ads shown as full screen images, and video ads displayed with 10 to 30 seconds skippable/unskippable video stream. The consumption of storage and network bandwidth of an ad varies significantly due to its format. We randomly pick one ad of each format from our test ad-network server. The size of each ad is listed in Table 6.5.

Note that the number of ads pre-downloaded per ad selection is set to 50 in Scenario 2. This means that, irrespective of format, 50 pieces of targeting attributes and URL information will need to be downloaded from the server and stored in the local cache. Thus, an extra storage and bandwidth of 8KB (0.16KB x 50) would be consumed.

At first glance, it might appear that 8KB is significant, when compared to only 0.7KB required by a raw text ad. However, a few factors should be kept in mind.

1. Raw text is rarely used alone for in-app ads. When compared to the consumption of more popular ad formats — image ads or video ads — the

consumption overhead only ranges from 1.4% to 28.5%. Note that, while we use compressed images and videos in the test ad-network server, ads in a real ad-server may be more resource-intensive. For example, on Google AdWords, the allowed max size of ad image is 150KB, and a skippable video ad could be up to 3 minutes long. In such cases, the consumption overhead could be further decreased.

2. From a user's perspective, 8KB per ad selection is bearable in the current 3G/4G/WIFI network environment. The consumption is also not significant for most mobile devices with relatively large storage capacity.

3. We assume that there are only slight changes in people's interests during a short period of time. Therefore, in a certain time frame, which is configurable, there is no necessity to download and store another new 8KB data for a new ad selection. The previously downloaded data can be reused when retrieving ads from the same app, or from different apps associated with the same ad-network in the period of time.

4. As reported by ThinkWithGoogle [145] in 2015, the average app user has 36 apps installed on their mobile, but only 9 are used daily. Another report, by MobileMarketing [146] in 2016, suggests that the average UK adult has 27 apps on their smartphone, but uses only 6 daily. Based on the data above, we assume that a user runs all the 36 installed apps every day in the worst case, and all apps are associated with different ad-networks. If the time frame of caching the list of ads is set to 5 days, then the finally consumption of storage and bandwidth of pre-downloading ads would be 1728KB per month, which is also tolerable.

Hence, taking into account the factors above, we would argue that the overall consumption of storage and bandwidth should not cause significant concerns on the user side.

Table 6.6: Consumption of memory and energy (PPTMA + AdSelector)

Term	cost
Memory consumption of PPTMA	16.47MB
Memory consumption of AdSelector	6.57MB
Energy consumption of PPTMA in standby mode for 24 hours	7.49mAh
Energy consumption of AdSelector in standby mode for 24 hours	5.62mAh
Totality of the Memory	3GB
Totality of the available energy (full-charged)	3220mAh

The final issue to be discussed is the memory and energy consumption of our solution. AdSelector, as an instance of the configurable ad-selection mechanism module that coordinates with PPTMA, must keep running in the background together with PPTMA since the startup of the mobile system. This fact indicates that it is unavoidable that certain impacts of performance on user experience would be introduced.

Related results of our simulation are shown in Table 6.6. The memory consumed by AdSelector and PPTMA is around 23MB in total, while the energy consumption of the whole framework which runs in the background for 24 hours only accounts for less than 1% of the battery. Therefore, in terms of the memory and energy consumption, the costs introduced by our proposed architecture would not notably degrade user experience considering the totality of the available resources.

6.4.5 Limitations and challenges

In this chapter, our focus has, somewhat inevitably, been on the users' perspective: that, after all, was the original motivation for the work. Many issues have yet to be explored from the perspectives of app developers and ad-networks. We outline initial thoughts in this regard now.

First, although AdSelector need not replace the existing infrastructure of ad-networks, slight modifications to business logic will be required. For example, new algorithms for selecting ads with coarse-grained user profiles need to be designed on the ad-networks' side, new workflows should be applied as the process of ad-selection

is divided into two stages, and the app developers might need to update their apps with new Ad-SDKs that could support the changes. The modification work in the implementation phase might limit the deployment of AdSelector.

Second, while on the users' side the time cost and consumption of storage and bandwidth is acceptable, the additional 8KB consumption and the extra network request for every single user will be a considerable overhead for an ad-network with millions of users.

Third, AdSelector provide an additional infrastructure — the trustworthy billing system — to provide the feature of reporting ad click without revealing involved users. Placing trust in a third-party billing system would also be a factor that would need to be overcome in any real implementation.

Additionally, in the existing targeted mobile advertising ecosystem, ad-networks are able to obtain a huge amount of fine-grained personal data. The user data is not only used in real-time targeting, but also used in other aspects such as designing marketing strategy or developing new advertising services. Applying privacy-preserving framework such as AdSelector would reduce their ability to collect users' personal information and consequently influence other businesses based on analysing fine-grained user data. Thus, provide ad-networks with enough incentives — for instance, lower users' hostility, higher response rates, and adapting to the trend of increasing privacy-consciousness — for deploying our framework is also a challenge.

6.5 Summary

In this chapter we have described a profile-based mobile advertisement selection mechanism, AdSelector, which aims to enable mobile users to take advantage of useful advertisements without disclosing their personal information. AdSelector is an instance of the privacy-aware ad selection mechanism that can be supported by PPTMA. The main functionality of AdSelector utilises the centralised management of personal information and fine-grained access control provided by PPTMA.

AdSelector consists of several mechanisms. The user subscription mechanism helps to increase user engagement; the two-stage ad selection process enables users

to obtain accurate ads based on their fine-grained profiles without disclosing the profiles to advertisers; the trustworthy billing system provides reliable ad-view and ad-click reports without revealing the involved users, and also assists in detecting advertising frauds. We have implemented an initial prototype for Android to simulate and evaluate AdSelector.

In conclusion, we believe that, given the continued growth of targeted mobile advertising, mechanisms such as the one described in this chapter will have an important role to play in terms of balancing the drivers of the economic model and the requirements of users in the emerging TMA ecosystem.

In the next chapter, we will discuss the roadmap for taking the work forward.

7
A Roadmap

Contents

7.1 Economic models of incentives 163
7.2 Potential application to other domains 166
7.3 Connection with wider academic areas 167
7.3.1 Internet privacy 168
7.3.2 Economics of privacy 168
7.4 Evolution of the TMA ecosystem 169
7.5 Summary 170

In previous chapters, we have described the current state of the TMA ecosystem, introduced our research question, and presented four main contributions pertaining to achieving the balance between privacy and utility in the context of TMA. Before we move forward to the conclusions, we first present some additional considerations that have emerged from our work. We start by discussing how our work might be taken forward, and, potentially, commercialised.

7.1 Economic models of incentives

Economic models and user incentives are important factors to be considered in real world environments. Hence, the first step in terms of making our work real, or,

Stakeholder	Inputs
User	Fine-grained personal data.
Advertiser	1. Advertising expenses based on ad click/view reports. 2. Losses caused by fraud attacks.
Ad-network	1. Expenses on personal data collection. 2. Expenses on users' preferences analysis. 3. Expenses on ads display on content providers' apps or websites. 4. Losses caused by fraud attacks.
Content provider	1. Resources of ads display. 2. Losses caused by fraud attacks.

Stakeholder	Outputs
User	Tailor-made ads based on passively submitted personal data.
Advertiser	1. Ad performance depending on the accuracy of ads-delivery. 2. Ad click/view reports provided by ad-networks.
Ad-network	1. Payment received from advertiser. 2. Dataset of users' personal information. 3. Hostility received from users.
Content provider	1. Payment shared by ad-networks based on ad click/view reports. 2. Ad click/view reports provided by ad-networks. 3. Hostility from users.

Table 7.1: Stakeholders' inputs and outputs without PPTMA

even better, commercialised, is to figure out how to motivate the stakeholders in the TMA area.

A brief comparison of each stakeholder's inputs and outputs is illustrated in Table 7.1 and Table 7.2 to show the differences that the PPTMA framework gives rise to. Through the comparison, incentives of applying the PPTMA framework can be clearly abstracted.

1. **Users** are able to enjoy higher quality tailor-made ads with only coarse-grained personal data being released. During this process, they can even decide which part of their personal data should be shared with which third-party. In addition, a mechanism for compensating privacy leakage, which will be introduced later in Section 7.3 as a wider academic link, could be applied to provide compensation for their exposed information, and to encourage users to get involved in the framework.

Stakeholder	Inputs
User	Coarse-grained personal data.
Advertiser	1. Advertising expenses based on ad click/view reports. 2. Less losses caused by fraud attacks.
Ad-network	1. Expenses on personal data collection. 2. Less expenses on users' preferences analysis due to the user subscription mechanism. 3. Expenses on ads display on content providers' apps or websites. 4. Less losses caused by fraud attacks. 5. Expenses as compensation to users for privacy leakage or ads clicks.
Content provider	1. Resources of ads display. 2. Less losses caused by fraud attacks.

Stakeholder	Outputs
User	More accurate tailor-made ads based on initiatively submitted personal data including user subscribed categories.
Advertiser	1. Ad performance depending on the more accurate delivery of ads. 2. More reliable ad click/view reports provided by ad-networks and verified by the trustworthy third-party PPTMA.
Ad-network	1. Payment received from advertiser. 2. Dataset of anonymous users' coarse-grained personal information. 3. More accurate preferences that initiatively submitted by anonymous users. 4. Less hostility received from users.
Content provider	1. Payment shared by ad-networks based on ad click/view reports. 2. More reliable ad click/view reports provided by ad-networks and verified by the trustworthy third-party PPTMA. 3. Less hostility from users.

Table 7.2: Changes imported by PPTMA

2. **Advertisers** are able to cut down the potential losses caused by fraud attacks as PPTMA provides related functions to detect fraud and verify the ad click/view reports. In addition, there is the potential to increase the ad performance as the targeting process would be more accurate due to the application of the user subscription mechanism.

3. **Ad-networks** are able to perform privacy-friendly targeting with PPTMA. Previous research [130, 131, 147] suggests that consumers' privacy concerns significantly influence their willingness to get involved in related economic activities. A framework like PPTMA has the potential to help ad-networks reduce users' hostility and uncertainty. As a result, more accurate ad clicks could be achieved.

4. **Content providers** once shared a part of users' hostility for ad-networks and bore the risk of ads fraud attacks. As described above, PPTMA is able to lighten the burden for them to a certain extent.

While the theoretical incentives of applying the PPTMA framework in the real world are clear, it is easy to foresee that large ad-networks will not give up their vested interests in the near future. However, as we argued in Chapter 5, the framework could still be attractive to individual advertisers and small ad-networks who hold limited resources. In the long run, such framework can help to establish a virtuous circle for the TMA ecosystem.

In terms of the potential to make the framework commercialised, possible economic models include charging payment from ad-networks for the saved computing resources and the improved advertising performance, or deducting a percentage of the compensation to privacy leakage from end users.

7.2 Potential application to other domains

In this section, we present some initial ideas of applying our work to other potential domains. We start by reviewing some core features of PPTMA:

1. PPTMA works as a trusted platform that provides centralised management facilities of users' privacy.

2. It provides a fine-grained access control mechanism for personal information, which allows users to interact with their own data on their devices.

3. It provides the ability to support users to obtain tailor-made online services with only coarse-grained personal data released to the service providers.

4. It enables commercial organisations to provide services in a privacy-friendly way.

5. It has the potential to be extended via the integration of new configurable plugins.

Such core features of PPTMA suggest some wider domains that the framework might be applied to.

An immediate domain is that of the recommendation system. The core concept of a recommendation system [148–150] is a mechanism that automatically recommends personalised information or items to a specific user based on her preferences. The TMA system is, to a certain extent, an instance of such recommendation systems. Hence, with appropriate adjustment, the function of recommending tailor-made ads provided by PPTMA can be adapted to other scenarios. Such a system might enable users to enjoy useful recommendations of news, books, music or other general items without releasing their fine-grained data. With the core features maintained, PPTMA could be extended to a Privacy-Preserving Recommendation Framework.

From the perspective of data management and access control, PPTMA has the potential to be integrated with mobile systems and then work as a centralised data and permission management module. Compared to the current permission mechanisms provided by Android and iOS, the ability of allowing users to maintain and submit different copies of user profiles according to different levels of granularity makes PPTMA a flexible solution. Hence, data management and access control on mobile devices might be another domain in which PPTMA might be applied.

Returning to the TMA area, there is also the potential to transfer PPTMA to a privacy-friendly ad-network itself. Due to the feature of the two-step ad selection process and the utilisation of accurate preferences that are initially submitted by users, more accurate ad selection can be achieved with less computing resources required from the servers of ad-network (as part of the computing process can be performed on the client-side). Thus, an ad-network with limited resources can be established based on PPTMA.

7.3 Connection with wider academic areas

Privacy issues pertaining to targeted mobile advertising, the research problem this dissertation focuses on, represents only a segment of the wider academic discourse pertaining to privacy. In this section, we briefly present two instances

to discuss the links between our work and two other areas: **Internet Privacy** and the **Economics of Privacy**.

7.3.1 Internet privacy

Due to the huge amount of personal information stored in mobile devices, as well as the large number of them, privacy concerns over the utilisation of mobile devices for gathering personal data emerge continuously. The work described in this dissertation was driven by these concerns.

Similar to mobile devices, there is another rapidly developing area that needs to be considered, and might be linked to our work — privacy-preservation in the Internet of Things (IoT).

According to Cisco's annual mobile data forecast in 2017 [151], the number of mobile-connected devices will reach 11.6 billion by 2021. Predictions regarding the number of IoT-connected devices can reach 36 to 46 billion by 2021 [152, 153], which is triple to fourfold the number of mobile-connected devices.

A Federal Trade Commission [154] report found that 150 million discrete data points can be generated by fewer than 10,000 households every day. An enormous amount of data will be generated in the future considering the number of IoT devices. Hence, IoT has the potential to become the next important privacy battlefield, which leaves users under the risk of targeting, eavesdropping, and other privacy-related attacks.

From the perspective of privacy-preservation, mechanisms similar to PPTMA might have a role to play.

7.3.2 Economics of privacy

The economic discussions of privacy is not new: the view that the protection of privacy might make economic activities inefficient, which is consistent with the privacy trade-off, appeared in the academic literature as early as the 1980s [155].

The economics of privacy has subsequently become a fast-moving discipline. In particular, the trade-offs associated with users data sharing and protection has became the natural realm of economics [156].

One suggestion is that users should be able to trade their personal data for benefits according to their own wishes, or be compensated when their information is compromised [30, 31, 157]. Such a model enables businesses to use monetary rewards to encourage users to provide their personal data [158].

Future technical solutions should not only support users by protecting their privacy, but should also provide the ability for users to utilise their own personal data. Further research into PPTMA could be connected with such activity to quantify the value of fine-grained personal information and to provide a tool for making the trade-off real.

7.4 Evolution of the TMA ecosystem

The research described in this dissertation started in late 2014. During the process of the D.Phil. research, there was staggering development in the TMA ecosystem and in the related market.

The number of smartphone users worldwide increased from 1.57 billion in 2014 to 2.32 billion in 2017 [159]. The three years have also seen a huge increase in the number of available apps in the Google Play Store — from 1.3 million to 3.3 million [160] — and a similar increase in the iTunes App Store — from 1.2 million to 3.1 million [161]. In addition, global mobile ad spending reached US $107.16 billion in 2017; it was only approximately US $29.06 billion in 2014 [162].

While the algorithms of user targeting and mechanisms of ad auction are revised and improved each passing day, new changes pertaining to the format of advertising are also imported by ad-networks continually. For example, Google officially updated the layout of AdWord in 2016 for better performance for advertisers. New technologies such as AR (Augmented Reality) was imported by Alibaba in 2016 to conduct theme marketing, and to improve customers' personalised

experiences. Facebook updated advertising policies several times to better exploit the potential of its targeting tools.

In addition, there are some new tendencies emerged in the field of TMA. The continuing increase of mobile users — the targeted advertising audiences — provides a strong guarantee for the sustainable development of the industry. Along with the recognition of TMA, consumers' knowledge of privacy risks also improves, which leads to the design of new privacy-preserving mechanisms in the field. Internet giants such as Google and Facebook, who hold most of the TMA resources, start to guide the direction and gradually lead the development of the ecosystem. New formats and technologies are constantly introduced to the field, and the price and spending of TMA will be further increased. In addition to the traditional stakeholders (users, advertisers, ad-networks, and content providers), governments and organisations begin to play important roles in monitoring and balancing the privacy trade-offs.

Overall, it is clear that TMA will continue to make great economic benefits in the near future. Additionally, this area will still be an active battlefield for privacy.

7.5 Summary

In this chapter we have discussed the evolved roadmap of our research problem. We presented the initial ideas of applying the framework in the real world environment. We discussed the potential to apply our work in wider domains and introduced its connection with other academic areas. We also briefly reviewed how the TMA ecosystem has evolved since the start of the D.Phil. research. In the next, and the final, chapter, we will examine the main contributions arising from our work, and suggest possible areas of future work.

8 Conclusion

Contents

8.1 A review 172
8.1.1 The user study: perspectives on privacy 172
8.1.2 Formal models 173
8.1.3 PPTMA 173
8.1.4 AdSelector 175
8.2 Measuring success against threats 176
8.3 Future work 177
8.3.1 Improve the framework 177
8.3.2 Improve and extend the formal models 178
8.3.3 Extend the framework to wider domains 178
8.3.4 Research TMA from another perspective 179
8.4 Outlook 179

This dissertation has explored the delicate balance between the goals of the advertisers and the consumers in order to address a research question:

> *Is it possible to develop a privacy-preserving TMA framework that enables mobile users to take advantage of useful advertising services without their privacy being compromised, and without impacting significantly advertising effectiveness?*

Driven by the research question, four contributions have been presented: (1) a user study investigating human factors involved in the privacy trade-offs; (2) the formal models that underpin our proposed privacy-preserving framework; (3)

the design and prototype implementation of the PPTMA framework; and (4) the design of AdSelector, a privacy-preserving advertisement selection mechanism that coordinates with PPTMA.

In this final chapter, we now provide a brief review of these contributions, suggest possible future work in this area, and then conclude.

8.1 A review

In this section, we briefly review the main points of the work described in this dissertation. The discussion of related contributions are organised according to the order in which they were introduced in our work.

8.1.1 The user study: perspectives on privacy

In Chapter 3, we discussed people's perspectives on privacy issues pertaining to mobile services based on the results of a user study. A list of high level requirements to address our research problem was also provided.

Surrounded with the unavoidable privacy trade-off and involved in the underlying free app ecosystem, people show various attitudes toward the related privacy issues. The user study saw great differences among the views of participants. The findings suggested that human factors such as users' awareness and knowledge of privacy risks, users' trust of service providers, users' desire for mobile services, and their belief of cyber privacy are all important contributing items in the decision-making process. Meanwhile, users are also encouraged to share their information in particular situations (e.g. when there is no substitute for the apps).

Based on the results of the user study, we discussed the goal, the high level requirements, and the threat models of a privacy-preserving system that has the potential to help users in their decision-making. In particular, high level requirements such as (1) privacy-preserving, (2) ad scanning and verifying, (3) controlled access, (4) data management, and (5) business supporting were then extended in the following chapters. These requirements finally formed a part of the core features

of the PPTMA framework (described in Chapter 5). The success of PPTMA will be measured against the threat models in Section 8.2.

8.1.2 Formal models

In Chapter 4, we demonstrated how formal models can help in analysing privacy in the context of TMA, in allowing users to specify control of their personal information, and in underpinning our proposed system.

First, important aspects of current TMA systems were abstracted and formalised with the schema language of Z. During the formalisation process, features of such systems that would impact users' privacy (e.g. collection of personal data and analysis of targeting) were identified.

Then we described a further specification that introduced three new subsystems: (1) *PermissionsSystem* that allows users to control how much of their information is released; (2) *LocalTMASystem* that allows the analysis of targeting and segmentation to be performed in users' clients; and (3)*BillingAssistantSystem* that helps to record click operations without exposing the users' information.

Enhanced with the new subsystems, the original TMA model was extended to a privacy-preserving TMA model, which underpinned the design of the PPTMA framework.

In addition, this work demonstrated how the balance between utility and privacy can be reasoned about with formal methods. It also illustrated the beneficial roles that can be played by formal models in the context of data sharing: the models can help users to better understand (and, to an extent, specify) the extent to which their personal information is shared.

8.1.3 PPTMA

In Chapter 5, the design of the prototype system PPTMA was introduced. The PPTMA framework sits at the core of our research: looking backward, PPTMA is proposed based on the results of the user study and the formal models; looking

forward, PPTMA provides the main framework for the integration of configurable ad-selection mechanisms such as AdSelector (described in Chapter 6).

The PPTMA system serves privacy-friendly targeted ads in two ways: via either a 'cooperative mode' or a 'mandatory mode'. In the cooperative mode, ad-networks are aware of PPTMA, and tailor-made ads are selected with users' local information on the client. In the mandatory mode, the system hooks API calls that can expose users' personal information, and consults a local policy to decide how to proceed.

At a high level, PPTMA works as a service that runs on mobile platforms to perform permission management and sensitive data processing between the underlying database of the mobile system and untrusted third-party applications. The following features are supported in the initial prototype.

1. **Centralised management of users' privacy.** PPTMA enables users to examine and customise their personal information through a unified interface. Different copies of the user's profile can be created and edited manually, then shared with different third-parties that have been approved by the user.

2. **Fine-grained access control.** This feature enables users to control permissions associated with accessing their personal information held by third-parties, and decide which parts of their personal information can be collected by which ad-networks. In the initial prototype of PPTMA, which runs on the Android system, this feature is implemented by hooking sensitive Android APIs.

3. **Ad-plugins scanning and verifying.** PPTMA enables users to discover ad-plugins embedded in the installed mobile apps and distinguish between the related ad-networks. Apps import ad libraries provided by ad-networks to perform targeted advertising. By scanning external libraries contained in apps and comparing the feature codes with pre-collected ad libraries, the system can identify which ad-networks are associated with a particular app, then deduce the ad styles and potential behaviour of the app. In addition, a behaviour monitoring service is also provided to detect malicious third-party apps.

4. **Support for privacy-aware advertisement selection.** PPTMA enables ad-networks to implement advertisement selection on mobile clients with customised algorithms. This feature enables ad-networks to create tailor-made ads without delivering users' personal information to the ad-networks' servers, which limits the use of users' personal information inside their mobile devices. To implement this feature, PPTMA makes use of the light version of ad-SDKs provided by cooperative ad-networks. The ad-SDKs enable PPTMA to perform some basic functions such as pre-downloading potential lists of ads and submitting trustworthy ad view/click reports to support local advertisement selection.

It is worth noting that, in addition to the TMA area, PPTMA also has the potential to be utilised in other domains (e.g. recommendation systems) as a centralised management tool to support user privacy.

8.1.4 AdSelector

In Chapter 6, we introduced AdSelector: a privacy-preserving advertisement selection mechanism for mobile devices. AdSelector is an instance of the privacy-aware ad selection mechanism that can be supported by PPTMA. The main functionality of AdSelector utilises the centralised management of personal information and fine-grained access control provided by PPTMA. In addition, the cooperative mode of PPTMA was used to examine the performance of AdSelector.

The main feature of AdSelector is its combination of several mechanisms. In particular: (1) the user subscription mechanism helps users to identify their interests and then increases user engagement; (2) the two-stage ad selection process helps users to enjoy accurate ads based on their fine-grained profiles — which are stored inside their local devices — and only provide coarse-grained information to ad servers; and (3) the trustworthy billing system provides reliable ad-click reports and assists in detecting advertising frauds without revealing the involved users.

AdSelector was designed to be implemented in two distinct ways: as a third-party app or as an external library. While there are some trade-offs associated

with both modes, the different implementations are helpful for the framework to be adapted to different environments.

8.2 Measuring success against threats

In Chapter 3, we presented the threat models of the traditional TMA system from the different perspectives of the stakeholders. Then, in Chapter 7, the changes to the TMA system that are imported by PPTMA were introduced. We now briefly measure the success of PPTMA against the threats faced by traditional TMA.

The main threats described in the threat models are summarised as follows:

1. To perform targeting, ad-networks would collect users' fine-grained personal information with ad-plugins.

2. Untrusted third-parties might abuse permissions to collect (and trade) users' personal data.

3. Users could block ads from particular ad-networks or apps, which influences the interests of advertisers, ad-networks and content providers.

4. Ad-networks and advertisers face the threat of click-fraud attacks performed by competitors or content providers. In addition, ad-networks could provide fake click/view reports to advertisers and content providers for economic benefit.

Correspondingly, the following improvements of PPTMA help to measure the success against the threats:

1. PPTMA enables users to decide the granularity of the personal data submitted to cooperative ad-networks. It also allows ad-networks to continue targeting in a privacy-friendly way.

2. PPTMA enables users to control permissions associated with each third-party app.

3. PPTMA has the potential to decrease hostility from users and provides more accurate tailor-made ads based on the user subscription mechanism. By doing this, the possibility of ads-blocking would be reduced.

4. PPTMA provides a trustworthy billing system to detect click-fraud attacks and verify the click/view reports. Thus, the more reliable click/view reports can be provided to ad-networks, advertisers and content providers respectively.

8.3 Future work

In this section, we consider some possible areas for future work which might follow from the work we described in this dissertation. The research concerns are organised into four groups: improve the framework, improve the formal models, extend the framework to wider domains, and research TMA from another perspective. We discuss them in turn.

8.3.1 Improve the framework

An immediate possible area of future research might be the improvement of the PPTMA framework itself.

In the TMA area, PPTMA provides an underpinning framework. AdSelector then provided an instance of the configurable ad-selection mechanism to support the framework. Improving the framework might involve the development and evaluation of particular algorithms for selecting ads on the basis of user preferences and permissions. In this respect, our prototype implementation has provided us with a starting point.

In addition, further testing and development of the prototype might also be a task. Given the fact that the initial prototypes of PPTMA and AdSelector are developed based on the Android system, an obvious follow-up opportunity would be to explore the possibility of transplanting the framework to other platforms such as iOS. The different architectures and mechanisms involved in these platforms

might be a new challenge. More work might be done in order to validate the performance of the framework on such platforms.

From the perspective of information transmission, the initial prototype of our framework mainly focuses on controlling information on the application level. To ensure that no private profile can be exposed in the communication between devices and ad-networks through the network level, the development of additional privacy-preserving protocols might also be another task.

Furthermore, there is also a potential to integrate additional mechanisms or systems with PPTMA to improve its performance. As an instance, researchers [163] have already started to investigate the possibility of combining PPTMA with a Federated Identities Identity Management (FIdM) system. The study demonstrates the potential of further enhancing the practicality and the user acceptance of PPTMA.

8.3.2 Improve and extend the formal models

Formal models can be beneficial in many ways. Our research illustrated how formal models can be helpful in analysing privacy in the context of TMA.

To further exploit the benefits, exploring means of refining the access control model by leveraging work on user-driven access control (see, for example, [164]) might be a possible contribution. Furthermore, as the prototype implementation might be further developed, transplanted and refined, the models of the framework could continue to be used to underpin model-based testing [165] in the future.

8.3.3 Extend the framework to wider domains

As discussed in Chapter 7, although the PPTMA framework was proposed and developed in the context of targeted advertising, it has the potential to be extended to other domains. For example, in addition to filtering and delivering targeted advertisements to users, the core feature of PPTMA also allows it to deliver information about music, books, news, social tags and products in general without releasing users' personal information. In other words, with appropriate adjustment, PPTMA can be utilised in a wide range of recommendation systems to protect

users' personal data. Hence, another possible research area might be to explore the possibility and to implement the adjustment.

Furthermore, there is also a potential to extend the framework to a centralised management tool that integrates with mobile systems. For example, the latest versions of Android and iOS both allow users to grant or revoke particular permissions on particular apps. However, the permission system is still a "give or not give" mechanism. Compared to such mechanisms, PPTMA allows users to decide how detailed information they want to release, which is more flexible for the users. Thus, an important possible area for future research might be to investigate the possibility of extending PPTMA to a system-level module.

8.3.4 Research TMA from another perspective

PPTMA was proposed with a view to explore the balance between the goals of the advertisers and the consumers: advertisers pursue profits by applying TMA, which violates consumers' privacy; consumers hope to benefit from useful mobile advertisements without compromising their personal information. As a result, PPTMA was implemented as a tool that should be deployed and managed by users, but has the potential to cooperate with privacy-friendly ad-networks. Hence, it is a solution that mostly works from the users' perspective.

It is important to clearly point out that consumers and ad-networks (and advertisers) are not forever enemies. To establish a better economic environment, ad-networks should also consider related privacy-preserving mechanisms to achieve a win-win situation. Thus, explore privacy-preserving technologies from the perspective of ad-networks might be another important area for future research.

8.4 Outlook

Given the continued growth of targeted mobile advertising and the rapid development of mobile technologies, it is easy to foresee that the mobile service providers of the future are going to collect an ever-increasing amount of personal data. Located in the centre of the privacy battlefield, users should be armed with enough

awareness and knowledge of privacy risks, as well as suitable data management technologies, to protect themselves.

However, technology has never been neutral. While the privacy-preserving technologies are adopted to help users, countless resources are spent by the service providers to develop new technologies to fight back — new technologies that can be used to exploit more benefits based on users' personal data. However, at a certain point in the future, all the stakeholders involved in the TMA ecosystem — users, advertisers, ad-networks, content providers, even governments — will need to realise that only by working together can they establish a healthy ecosystem that can be sustainably developed.

We should try our best to prevent the all-too-commonly uttered "Nothing is private online" from becoming a solid reality.

Appendices

A

Questionnaire used in the qualitative study

1. Age:
   (a) 18 — 25
   (b) 26 — 35
   (c) 36 — 45
   (d) Over 45
2. Gender:
   (a) Male
   (b) Female
3. Would you say you live in:
   (a) Urban
   (b) Suburban
   (c) Rural

4. What is your highest level of education?

   (a) Primary education

   (b) Secondary education

   (c) Post-secondary

   (d) Upper secondary

   (e) Tertiary

   (f) Post-graduate

5. What is your occupation?

   (a) Professional and managerial occupation

   (b) Supervisory and technical occupation

   (c) Other white collar worker

   (d) Semi-skilled worker

   (e) Manual worker

   (f) Student

   (g) House person

   (h) Retired

   (i) Other

6. For which purposes do you use your phone? Please choose as many as you like.

   (a) To search for information (e.g. news, weather forecast, maps)

   (b) To make use of e-services (e.g. mobile bank)

   (c) To create and check emails

   (d) Online shopping

   (e) Social networking

(f) To play mobile games

(g) To read e-books

(h) To take photos or videos

(i) To play music

(j) Diary management (schedule, calendar)

(k) Other

7. What is the operating system of your phone?

(a) Android

(b) iOS

(c) Windows

(d) Other

8. To the best of your knowledge, which of the following information is technically possible to be collected/deduced by an app provider when you use an app? Please choose as many as you like.

(a) Time of usage

(b) Your geographical location

(c) Your email address

(d) Your phone number

(e) Your SMSs

(f) Your contacts

(g) Your calling history

(h) Your network operator

(i) Your IMEI number (IMEI number is a 15 digit unique ID of the mobile phone)

(j) The operating system of your phone

(k) The screen size of your phone

(l) The retail price of your phone

(m) The events of your calendar

(n) The list of apps installed on your phone

(o) The time and frequency of your app usage

(p) Your age

(q) Your gender

(r) Your interests

(s) None of these

(t) Don't know

(u) Other

9. In your opinion, what information should not be collected by any app, no matter the conditions? Please choose as many as you like.

(a) Time of usage

(b) Your geographical location

(c) Your email address

(d) Your phone number

(e) Your SMSs

(f) Your contacts

(g) Your calling history

(h) Your network operator

(i) Your IMEI number (IMEI number is a 15 digit unique ID of the mobile phone)

(j) The operating system of your phone

(k) The screen size of your phone

(l) The retail price of your phone

(m) The events of your calendar

(n) The list of apps installed on your phone

(o) The time and frequency of your app usage

(p) Your age

(q) Your gender

(r) Your interests

(s) None of these

(t) Don't know

(u) Other

10. For the options you didn't check in the last question, what might encourage you to give the information to an app? Please choose as many as you like.

(a) The app provides a statement regarding how the information is going to be used

(b) In exchange for access to the main functions of the app (e.g. giving location information to a weather forecast app for more accurate weather information)

(c) In exchange for valued-added service (e.g. giving demographic information to convert an app from trial version to full-function version)

(d) In exchange for a small discount or coupon

(e) In exchange for a useful recommendation (e.g. recommendation of local restaurants, interesting products, events, etc.)

(f) If the data would only be used in aggregate form

(g) If there is no substitute for the app (App store)

(h) Other

Open-ended questions:

1. Supposing you are posting stories/photos on Facebook or Instagram... Please answer the following questions:

   (a) What personal information is being collected?

   (b) Who is collecting your information? Where will the information end up?

   (c) How do they collect your information?

   (d) Why do they collect your information? Do you see any risks?

   (e) So you are exchanging your personal information for the utility provided by Facebook or instagram. Is it worth the cost?

2. Sometimes you need to tell apps your fine-grained information to obtain their services. Supposing there is a kind of privacy-preserving technologies that allows you to obtain the same services by giving apps only coarse-grained information. For example:

   Fine-grained information:
   Gender: Male
   Age: 28
   Location: Oxford, UK
   Interests: Basketball

   Coarse-grained information:
   Gender: Not available
   Age: 25-40
   Location: UK
   Interests: Team Sports (football, basketball, baseball)

Do such technologies influence your decision making progress when you face the trade-off between privacy and utility in mobile services? Please can you give an example?

3. In a restaurant, the cook needs a certificate and the food quality needs to be checked by government department. In terms of the privacy-preserving technologies, who do you think is responsible to certify such technologies do what they say (e.g. government department, organisations, famous Companies or others)?

B

Slides used in focus group sessions of the qualitative study

Page 0, Outline:

1. Introduction: 5 mins
2. A card-exchanging game: 10 mins.
3. Everyday experiences: 40 mins.
4. An example of privacy-preserving technologies: 5 mins.
5. Discussion: Benefits and barriers of such technologies: 20 mins.
6. Brief summary: 5 mins.

Slides:

1. Introduction:
   (a) Purpose.
   (b) Duration.
   (c) Confidentiality.

2. Ground rules:
   (a) Respect each other's opinions.
   (b) Respect each other's confidentiality.
   (c) Mobile phones on silent.
   (d) There is no right or wrong answers. Just tell us the first thing that pops into your mind on each topic.
   (e) Introduction of the observer.
3. Card game! (Introduction and examples.)
   (a) Personal Information versus Mobile Services.
   (b) Privacy versus Utility.
   (c) Cost versus Benefit.
4. Card game! (Think about the follows.)
   (a) Clear idea of the trade-off?
   (b) First round versus Second round: More services or less? Is it worth the cost?
   (c) First round versus Second round: More information or less? Which strategy do your prefer?
   (d) Involved factors?
5. Card game! (Mental associations.)
   (a) There exists the privacy trade-off.
   (b) Many factors are involved.

6. Everyday experiences: for each scenario, please think about and comment on the following topics:

   (a) What type of personal information do you think is being collected in this scenario?

   (b) Who do you think is collecting your personal information? Where do you think your personal information will end up?

   (c) How do you think they collect or deduce your information?

   (d) In what ways do you think your personal information will be used? Do you see any potential risks?

   (e) What utility do you think is obtained in this scenario? Do you think it is worth the cost?

7. Related scenarios:

   (a) Experience of exposing personal information actively: imagine that you are posting stories or photos on Facebook or Instagram.

   (b) Experience of interacting with mobile system providers: imagine that you are downloading an app from the official Android or iOS app stores.

   (c) Experience of receiving targeted recommendation services: imagine that you are visiting an unfamiliar place and you receive recommendations for useful information such as tourist attractions, hotels and restaurants on your mobile device.

   (d) Experience related to financial activities: imagine that you are doing online shopping on your mobile device.

   (e) Experience related to personal events: imagine that you are managing calendar events on your mobile device.

8. An example of privacy-preserving technologies. Introduction of a kind of privacy-preserving technologies that allows users to obtain the same services

by giving apps only coarse-grained information. For example:

Fine-grained information:
Gender: Male
Age: 28
Location: Oxford, UK
Interests: Basketball

Coarse-grained information:
Gender: Not available
Age: 25-40
Location: UK
Interests: Team Sports (football, basketball, baseball)

Such privacy-preserving technologies enable you to decide:

(a) Which part of your personal information is submitted to the service providers.

(b) Which third-party can access to the information you submitted.

(c) The granularity of the information you submitted.

9. Discussion: Benefits and barriers of such technologies.

(a) What do you see as the benefits of using such privacy-preserving technologies?

(b) What do you see as the barriers to using such privacy-preserving technologies?

(c) Do you think such privacy-preserving technologies could influence your decision-making process of making trade-offs between privacy and utility in mobile services?

(d) Who do you think should be responsible to certify such privacy-preserving technologies do what they say?

(e) (optional) What other mobile services do you think can also adopt such privacy-preserving technologies?

(f) (optional) What would you consider as an acceptable granularity of submitted information? And the acceptable length of storage on the servers?

10. Brief summary.

References

[1] Nigel Hollis. "Ten years of learning on how online advertising builds brands". In: *Journal of advertising research* 45.2 (2005), pp. 255–268.

[2] Agatha M Cole. "Internet Advertising After Sorrell v. IMS Health: A Discussion on Data Privacy & the First Amendment". In: *Cardozo Arts & Entertainment Law Journal* 30 (2012), pp. 283–315.

[3] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas. "Adnostic: Privacy preserving targeted advertising." In: *Proceedings of the 17th Annual Network and Distributed System Security Symposium (NDSS 2010)*. Retrieved April 6, 2016 from `https://www.isoc.org/isoc/conferences/ndss/10/pdf/05.pdf`. San Diego, CA, USA, 2010.

[4] Jun Yan, Ning Liu, Gang Wang, Wen Zhang, Yun Jiang, and Zheng Chen. "How much can behavioral targeting help online advertising?" In: *Proceedings of the 18th International Conference on World Wide Web (WWW 2009)*. ACM. Madrid, Spain, 2009, pp. 261–270.

[5] Howard Beales. *The value of behavioral targeting*. `http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf`. [Last accessed April 2015]. 2010.

[6] Avi Goldfarb and Catherine E Tucker. "Privacy regulation and online advertising". In: *Management Science* 57.1 (2011), pp. 57–71.

[7] R Bender. *Mobile-ad market still faces hurdles*. Wall Street Journal, February 18th, 2001. [Last accessed March 2016]. 2011.

[8] eMarketer. *US digital display ad spending to surpass search ad spending in 2016*. `http://www.emarketer.com/Article/US-Digital-Display-Ad-Spending-Surpass-Search-Ad-Spending-2016/1013442`. [Last accessed June 2016]. 2016.

[9] eMarketer. *Mobile is driving UK ad spend growth*. `http://www.emarketer.com/Article/Mobile-Driving-UK-Ad-Spend-Growth/1013685`. [Last accessed June 2016]. 2016.

[10] Ilias Leontiadis, Christos Efstratiou, Marco Picone, and Cecilia Mascolo. "Don'T Kill My Ads!: Balancing Privacy in an Ad-supported Mobile Application Market". In: *Proceedings of the 13th Workshop on Mobile Computing Systems & Applications (HotMobile 2012)*. San Diego, California: ACM, New York, 2012, 2:1–2:6.

[11] Yang Liu and A.C. Simpson. "Privacy-preserving targeted mobile advertising: Requirements, design and a prototype implementation". In: *Software: Practice and Experience* 46.12 (2016), pp. 1657–1684.

[12] Melody M Tsang, Shu-Chun Ho, and Ting-Peng Liang. "Consumer attitudes toward mobile advertising: An empirical study". In: *International Journal of Electronic Commerce* 8.3 (2004), pp. 65–78.

[13] Marko Merisavo, Sami Kajalo, Heikki Karjaluoto, Ville Virtanen, Sami Salmenkivi, Mika Raulas, and Matti Leppäniemi. "An empirical study of the drivers of consumer acceptance of mobile advertising". In: *Journal of Interactive Advertising* 7.2 (2007), pp. 41–50.

[14] Hamed Haddadi, Pan Hui, and Ian Brown. "MobiAd: Private and Scalable Mobile Advertising". In: *Proceedings of the 5th ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2010)*. Chicago, Illinois, USA: ACM, New York, 2010, pp. 33–38.

[15] H. Ahn, K.-J. Kim, and I. Han. "Mobile advertisement recommender system using collaborative filtering: MAR-CF". In: *Proceedings of the 2006 Conference of the Korea Society of Management Information Systems*. The Korea Society of Management Information Systems, 2006, pp. 709–715.

[16] Soe-Tsyr Yuan and You Wen Tsao. "A recommendation mechanism for contextualized mobile advertising". In: *Expert Systems with Applications* 24.4 (2003), pp. 399–414.

[17] Evelyne Beatrix Cleff. "Privacy Issues in Mobile Advertising". In: *International Review of Law, Computers & Technology* 21.3 (2007), pp. 225–236. eprint: `http://dx.doi.org/10.1080/13600860701701421`.

[18] Linfeng Zhang and Yong Guan. "Detecting click fraud in pay-per-click streams of online advertising networks". In: *The 28th International Conference on Distributed Computing Systems (ICDCS 2008)*. IEEE. Beijing, China, 2008, pp. 77–84.

[19] Alessandro Acquisti and Jens Grossklags. "Privacy and rationality in individual decision making". In: *IEEE Security & Privacy* 3.1 (2005), pp. 26–33.

[20] Myslewski, Rik. *Apple ads to target your iTunes history.* `http://www.theregister.co.uk/Print/2010/07/06/apple_targets_ads/`. [Last accessed April 2015].

[21] Pinjia Zou, Yanyan Xu, Zheng Fang, and Wei Li. "The effectiveness of location-based advertising: When, where, and to whom". In: *International Journal of Mobile Communications* 14.3 (2016), pp. 273–290.

[22] Byoungjip Kim, Jin-Young Ha, SangJeong Lee, Seungwoo Kang, Youngki Lee, Yunseok Rhee, Lama Nachman, and Junehwa Song. "AdNext: A Visit-pattern-aware Mobile Advertising System for Urban Commercial Complexes". In: *Proceedings of the 12th Workshop on Mobile Computing Systems and Applications (HotMobile 2011)*. Phoenix, Arizona, USA: ACM New York, 2011, pp. 7–12.

[23] G Bilchev and D Marston. "Personalised advertising-exploiting the distributed user profile". In: *BT Technology Journal* 21.1 (2003), pp. 84–90.

[24] Saikat Guha, Bin Cheng, and Paul Francis. "Privad: Practical Privacy in Online Advertising." In: *Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation (NSDI 2011)*. Boston, MA, USA: ACM, New York, 2011, pp. 169–182.

[25] Wenhao Li, Haibo Li, Haibo Chen, and Yubin Xia. "Adattester: Secure online mobile advertisement attestation using trustzone". In: *Proceedings of the 13th Annual International Conference on Mobile Systems, Applications, and Services*. ACM. 2015, pp. 75–88.

[26] Bin Liu, Suman Nath, Ramesh Govindan, and Jie Liu. "DECAF: Detecting and Characterizing Ad Fraud in Mobile Apps". In: *11th USENIX Symposium on Networked Systems Design and Implementation (NSDI 14)*. Seattle, WA: USENIX Association, 2014, pp. 57–70.

[27] Jonathan Crussell, Ryan Stevens, and Hao Chen. "Madfraud: Investigating ad fraud in android applications". In: *Proceedings of the 12th annual international conference on Mobile systems, applications, and services*. ACM. 2014, pp. 123–134.

[28] Hamed Haddadi. "Fighting online click-fraud using bluff ads". In: *ACM SIGCOMM Computer Communication Review* 40.2 (2010), pp. 21–25.

[29] Lichen Zhang, Xiaoming Wang, Junling Lu, Peng Li, and Zhipeng Cai. "An efficient privacy preserving data aggregation approach for mobile sensing". In: *Security and Communication Networks* 9.16 (2016), pp. 3844–3853.

[30] Kobbi Nissim, Claudio Orlandi, and Rann Smorodinsky. "Privacy-aware Mechanism Design". In: *Proceedings of the 13th ACM Conference on Electronic Commerce (EC 2012)*. Valencia, Spain: ACM, New York, 2012, pp. 774–789.

[31] Wei Wang, Linlin Yang, Yanjiao Chen, and Qian Zhang. "A privacy-aware framework for targeted advertising". In: *Computer Networks* 79 (2015), pp. 17–29.

[32] Saikat Guha, Alexey Reznichenko, Kevin Tang, Hamed Haddadi, and Paul Francis. "Serving Ads from localhost for Performance, Privacy, and Profit." In: *Proceedings of Hot Topics in Networking (HotNets 2009)*. 2009.

[33] Hamed Haddadi, Pan Hui, Tristan Henderson, and Ian Brown. "Targeted advertising on the handset: Privacy and security challenges". In: *Pervasive Advertising*. Springer, 2011, pp. 119–137.

[34] Matthew Fredrikson and Benjamin Livshits. "Repriv: Re-imagining content personalization and in-browser privacy". In: *Proceedings of IEEE Symposium on Security and Privacy (SP 2011)*. 2011, pp. 131–146.

[35] Michaela Hardt and Suman Nath. "Privacy-aware personalization for mobile advertising". In: *Proceedings of the 2012 ACM conference on Computer and communications security (CCS 2012)*. ACM. Raleigh, NC, USA, 2012, pp. 662–673.

[36] Amin Tootoonchian, Stefan Saroiu, Yashar Ganjali, and Alec Wolman. "Lockr: better privacy for social networks". In: *Proceedings of the 5th International Conference on Emerging Networking Experiments and Technologies*. ACM. 2009, pp. 169–180.

[37] Ramón Cáceres, Landon Cox, Harold Lim, Amre Shakimov, and Alexander Varshavsky. "Virtual individual servers as privacy-preserving proxies for mobile devices". In: *Proceedings of the 1st ACM Workshop on Networking, Systems, and Applications for Mobile Handhelds*. ACM. 2009, pp. 37–42.

[38] Saikat Guha, Kevin Tang, and Paul Francis. “NOYB: Privacy in online social networks”. In: *Proceedings of the 1st Workshop on Online Social Networks*. ACM. 2008, pp. 49–54.

[39] Daniel C Howe and Helen Nissenbaum. “TrackMeNot: Resisting surveillance in web search”. In: *Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society* 23 (2009), pp. 417–436.

[40] Imdad Ullah, Roksana Boreli, Salil S Kanhere, and Sanjay Chawla. “ProfileGuard: Privacy Preserving Obfuscation for Mobile User Profiles”. In: *Proceedings of the 13th Workshop on Privacy in the Electronic Society (WPES 2014)*. ACM. 2014, pp. 83–92.

[41] 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 (HotMobile 2011)*. ACM. 2011, pp. 49–54.

[42] Mike Ter Louw, Karthik Thotta Ganesh, and VN Venkatakrishnan. “AdJail: Practical Enforcement of Confidentiality and Integrity Policies on Web Advertisements”. In: *USENIX Security Symposium*. 2010, pp. 371–388.

[43] Xinshu Dong, Minh Tran, Zhenkai Liang, and Xuxian Jiang. “AdSentry: Comprehensive and flexible confinement of JavaScript-based advertisements”. In: *Proceedings of the 27th Annual Computer Security Applications Conference*. ACM. 2011, pp. 297–306.

[44] William Enck, Peter Gilbert, Seungyeop Han, Vasant Tendulkar, Byung-Gon Chun, Landon P Cox, Jaeyeon Jung, Patrick McDaniel, and Anmol N Sheth. “TaintDroid: An information-flow tracking system for realtime privacy monitoring on smartphones”. In: *ACM Transactions on Computer Systems (TOCS)* 32.2 (2014), p. 5.

[45] Anthony Desnos and Patrik Lantz. *Droidbox: An Android application sandbox for dynamic analysis*. `http://www.honeynet.org/gsoc2011/slot5`. [Last accessed April 2015].

[46] Lok-Kwong Yan and Heng Yin. “DroidScope: Seamlessly Reconstructing the OS and Dalvik Semantic Views for Dynamic Android Malware Analysis.” In: *Proceedings of the 21th USENIX Security Symposium (USENIX Security 2012)*. 2012, pp. 569–584.

[47] Michael Spreitzenbarth, Felix Freiling, Florian Echtler, Thomas Schreck, and Johannes Hoffmann. “Mobile-sandbox: Having a deeper look into Android applications”. In: *Proceedings of the 28th Annual ACM Symposium on Applied Computing (SAC 2013)*. ACM. 2013, pp. 1808–1815.

[48] Ming Li, Ning Cao, Shucheng Yu, and Wenjing Lou. “Findu: Privacy-preserving personal profile matching in mobile social networks”. In: *INFOCOM*. IEEE. 2011, pp. 2435–2443.

[49] Leucio Antonio Cutillo, Refik Molva, and Melek Onen. “Safebook: A distributed privacy preserving online social network”. In: *IEEE International Symposium on World of Wireless, Mobile and Multimedia Networks (WoWMoM)*. IEEE. 2011, pp. 1–3.

[50] Leudo Antonio Cutillo, Refik Molva, and Thorsten Strufe. "Privacy preserving social networking through decentralization". In: *6th International Conference on Wireless On-Demand Network Systems and Services*. IEEE. 2009, pp. 145–152.

[51] Marco Voss, Andreas Heinemann, and Max Muhlhauser. "A privacy preserving reputation system for mobile information dissemination networks". In: *1st International Conference on Security and Privacy for Emerging Areas in Communications Networks*. IEEE. 2005, pp. 171–181.

[52] Ari Juels. "Targeted advertising... and privacy too". In: *Topics in Cryptology—CT-RSA 2001*. Springer, 2001, pp. 408–424.

[53] Michael Backes, Aniket Kate, Matteo Maffei, and Kim Pecina. "Obliviad: Provably secure and practical online behavioral advertising". In: *IEEE Symposium on Security and Privacy*. IEEE. 2012, pp. 257–271.

[54] Appthority. *Appthority exposes security and privacy risks behind top 400 mobile apps*. `https://www.appthority.com/company/news-and-events/press-releases/appthority-exposes-security-and-privacy-risks-behind-top-400-mobile-apps/`. [Last accessed July 2017]. 2014.

[55] Pew Research Center. *Privacy and data management on mobile devices*. `http://www.pewinternet.org/2012/09/05/privacy-and-data-management-on-mobile-devices/`. [Last accessed July 2017]. 2012.

[56] Alastair R Beresford, Dorothea Kübler, and Sören Preibusch. "Unwillingness to pay for privacy: A field experiment". In: *Economics Letters* 117.1 (2012), pp. 25–27.

[57] Statista. *Daily time spent online via mobile 2016, by generation*. `https://www.statista.com/statistics/428425/daily-time-spent-online-mobile-age/`. [Last accessed July 2017]. 2016.

[58] Statista. *Number of apps available in leading app stores 2016*. `https://www.statista.com/statistics/276623/number-of-apps-available-in-leading-app-stores/`. [Last accessed December 2016]. 2016.

[59] AppBrain. *Google Play Stats*. `http://www.appbrain.com/stats`. [Last accessed December 2017]. 2017.

[60] Jinyan Zang, Krysta Dummit, James Graves, Paul Lisker, and Latanya Sweeney. "Who knows what about me? A survey of behind the scenes personal data sharing to third parties by mobile apps". In: *Technology Science* (2015).

[61] Ginny Marvin. *Survey: 3 out of 4 consumers now notice retargeted ads*. `http://marketingland.com/3-out-4-consumers-notice-retargeted-ads-67813`. [Last accessed July 2017]. 2016.

[62] Fen Liu, Xuefeng Zhao, Patrick Y.K. Chau, and Qing Tang. "Roles of perceived value and individual differences in the acceptance of mobile coupon applications". In: *Internet Research* 25.3 (June 2015), pp. 471–495.

[63] Heng Xu, Hock-Hai Teo, Bernard CY Tan, and Ritu Agarwal. "The role of push-pull technology in privacy calculus: The case of location-based services". In: *Journal of Management Information Systems* 26.3 (2009), pp. 135–174.

[64] Iris A Junglas, Norman A Johnson, and Christiane Spitzmüller. "Personality traits and concern for privacy: an empirical study in the context of location-based services". In: *European Journal of Information Systems* 17.4 (2008), pp. 387–402.

[65] Susan B Barnes. "A privacy paradox: Social networking in the United States". In: *First Monday* 11.9 (2006).

[66] Naveen Farag Awad and MS Krishnan. "The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization". In: *MIS Quarterly* 30.1 (2006), pp. 13–28.

[67] Patricia A Norberg, Daniel R Horne, and David A Horne. "The privacy paradox: Personal information disclosure intentions versus behaviors". In: *Journal of Consumer Affairs* 41.1 (2007), pp. 100–126.

[68] Sonja Utz and Nicole Krämer. "The privacy paradox on social network sites revisited: The role of individual characteristics and group norms". In: *Cyberpsychology: Journal of Psychosocial Research on Cyberspace* 3.2 (2009), p. 2.

[69] Bernhard Debatin, Jennette P Lovejoy, Ann-Kathrin Horn, and Brittany N Hughes. "Facebook and online privacy: Attitudes, behaviors, and unintended consequences". In: *Journal of Computer-Mediated Communication* 15.1 (2009), pp. 83–108.

[70] Kirsten Martin. "Privacy Notices as Tabula Rasa: An empirical investigation into how complying with a privacy notice is related to meeting privacy expectations online." In: *Journal of Public Policy & Marketing* (2014).

[71] Ryan Calo. "Against notice skepticism in privacy (and elsewhere)". In: *Notre Dame Law Review* 87 (2013), pp. 1027–1073.

[72] George R Milne and Mary J Culnan. "Strategies for reducing online privacy risks: Why consumers read (or don't read) online privacy notices". In: *Journal of Interactive Marketing* 18.3 (2004), pp. 15–29.

[73] Catherine E Tucker. "The economics of advertising and privacy". In: *International Journal of Industrial Organization* 30.3 (2012), pp. 326–329.

[74] Catherine E Tucker. "Social networks, personalized advertising and privacy controls". In: *Journal of Marketing Research* 51.5 (2014), pp. 546–562.

[75] Mary Madden, Sandra Cortesi, Urs Gasser, Amanda Lenhart, and Maeve Duggan. *Parents, Teens and Online Privacy*. `http://www.pewinternet.org/2012/11/20/parents-teens-and-online-privacy/`. [Last accessed April 2015].

[76] Mary Madden, Amanda Lenhart, Sandra Cortesi, Urs Gasser, Maeve Duggan, Arron Smith, and Meredith Beaton. *Teens, Social Media, and Privacy*. `http://www.pewinternet.org/2013/05/21/teens-social-media-and-privacy/`. [Last accessed April 2015].

[77] Mary Madden. *Public Perceptions of Privacy and Security in the Post-Snowden Era*. `http://www.pewinternet.org/2014/11/12/public-privacy-perceptions/`. [Last accessed April 2015].

[78] Alice E Marwick. ""I'm a lot more interesting than a Friendster profile": Identity presentation, authenticity, and power in social networking services". In: *Association of Internet Researchers* 6 (2005).

[79] Zeynep Tufekci. "Can you see me now? Audience and disclosure regulation in online social network sites". In: *Bulletin of Science, Technology & Society* 28.1 (2008), pp. 20–36.

[80] 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 18th Symposium on Usable Privacy and Security (SOUPS 2012)*. ACM. 2012, p. 3.

[81] Michael Kern and Johannes Sametinger. "Permission tracking in Android". In: *Proceedings of the 6th International Conference on Mobile Ubiquitous Computing, Systems, Services and Technologies (UBICOMM 2012)*. Barcelona, Spain, 2012, pp. 148–155.

[82] Shashank Holavanalli, Don Manuel, Vishwas Nanjundaswamy, Brian Rosenberg, Feng Shen, Steven Y Ko, and Lukasz Ziarek. "Flow permissions for Android". In: *Proceedings of the 28th IEEE/ACM International Conference on Automated Software Engineering (ASE 2013)*. IEEE. Palo Alto, CA, USA, 2013, pp. 652–657.

[83] Mohammad Nauman, Sohail Khan, and Xinwen Zhang. "Apex: Extending Android permission model and enforcement with user-defined runtime constraints". In: *Proceedings of the 5th ACM Symposium on Information, Computer and Communications Security (ASIACCS 2010)*. ACM. Beijing, China, 2010, pp. 328–332.

[84] Google. *Android platform versions*. `https://developer.android.com/about/dashboards/index.html`. [Last accessed March 2017].

[85] Kim Bartel Sheehan. "Toward a typology of Internet users and online privacy concerns". In: *The Information Society* 18.1 (2002), pp. 21–32.

[86] Asunción Esteve. "The business of personal data: Google, Facebook, and privacy issues in the EU and the USA". In: *International Data Privacy Law* 7.1 (2017), pp. 36–47.

[87] Phillipa Gill, Vijay Erramilli, Augustin Chaintreau, Balachander Krishnamurthy, Konstantina Papagiannaki, and Pablo Rodriguez. "Follow the money: understanding economics of online aggregation and advertising". In: *Proceedings of the 2013 conference on Internet measurement conference (IMC 2013)*. ACM. 2013, pp. 141–148.

[88] Bernadette Kamleitner, Stephan Dickert, Marjan Falahrastegar, and Hamed Haddadi. "Information bazaar: a contextual evaluation". In: *Proceedings of the 5th ACM Workshop on Hotplanet (Hotplanet 2013)*. ACM. 2013, pp. 57–62.

[89] Sonam Samat and Alessandro Acquisti. "Format vs. Content: The Impact of Risk and Presentation on Disclosure Decisions". In: *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS 2017)*. 2017, pp. 377–384.

[90] Bram Bonné, Sai Teja Peddinti, Igor Bilogrevic, and Nina Taft. "Exploring decision making with Android's runtime permission dialogs using in-context surveys". In: *Proceedings of the 13th Symposium on Usable Privacy and Security (SOUPS 2017)*. 2017.

[91] Jenny Kitzinger. "Qualitative research. Introducing focus groups." In: *BMJ: British Medical Journal* 311.7000 (1995), p. 299.

[92] Raymond L Gorden. *Interviewing: Strategy, techniques, and tactics*. Dorsey press Homewood, IL, 1969.

[93] Rick Wash, Emilee Rader, and Chris Fennell. "Can people self-report security accurately?: Agreement between self-report and behavioral measures". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM. 2017, pp. 2228–2232.

[94] Nadia Olivero and Peter Lunt. "Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control". In: *Journal of Economic Psychology* 25.2 (2004), pp. 243–262.

[95] Stefanie Pötzsch. "Privacy awareness: A means to solve the privacy paradox?" In: *IFIP Summer School on the Future of Identity in the Information Society*. Springer. 2008, pp. 226–236.

[96] Curt J Dommeyer and Barbara L Gross. "What consumers know and what they do: An investigation of consumer knowledge, awareness, and use of privacy protection strategies". In: *Journal of Interactive Marketing* 17.2 (2003), pp. 34–51.

[97] Serge Egelman, Marian Harbach, and Eyal Peer. "Behavior ever follows intention?: A validation of the security behavior intentions scale (SeBIS)". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*. CHI '16. Santa Clara, California, USA: ACM, 2016, pp. 5257–5261.

[98] 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 9th Symposium on Usable Privacy and Security (SOUPS 2013)*. ACM. 2013, p. 12.

[99] Neelima Sailaja, Andy Crabtree, and Phil Stenton. "Challenges of using personal data to drive personalised electronic programme guides". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM. 2017, pp. 5226–5231.

[100] Charles Comegys, Mika Hannula, and Jaaui Váisánen. "Effects of consumer trust and risk on online purchase decision-making: A comparison of Finnish and United States students". In: *International Journal of Management* 26.2 (2009), p. 295.

[101] Heng Xu and Sumeet Gupta. "The effects of privacy concerns and personal innovativeness on potential and experienced customers' adoption of location-based services". In: *Electronic Markets* 19.2-3 (2009), pp. 137–149.

[102] Shuk Ying Ho and Sai Ho Kwok. "The attraction of personalized service for users in mobile commerce: An empirical study". In: *ACM SIGecom Exchanges* 3.4 (2002), pp. 10–18.

[103] 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". In: *Personal and Ubiquitous Computing* 13.6 (2009), pp. 401–412.

[104] Chanda Phelan, Cliff Lampe, and Paul Resnick. "It's creepy, but it doesn't bother me". In: *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems (CHI '16)*. ACM. 2016, pp. 5240–5251.

[105] Daniel J Solove. "I've got nothing to hide and other misunderstandings of privacy". In: *San Diego Law Review* 44 (2007), p. 745.

[106] Max Van Kleek, Ilaria Liccardi, Reuben Binns, Jun Zhao, Daniel J Weitzner, and Nigel Shadbolt. "Better the devil you know: Exposing the data sharing practices of smartphone apps". In: *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems (CHI '17)*. ACM. 2017, pp. 5208–5220.

[107] Julia B Earp and David Baumer. "Innovative web use to learn about consumer behavior and online privacy". In: *Communications of the ACM* 46.4 (2003), pp. 81–83.

[108] Vincent F. Taylor and Ivan Martinovic. "SecuRank: Starving Permission-Hungry Apps Using Contextual Permission Analysis". In: *Proceedings of the 6th Workshop on Security and Privacy in Smartphones and Mobile Devices*. SPSM '16. Vienna, Austria: ACM, 2016, pp. 43–52.

[109] Zinaida Benenson, Anna Girard, and Ioannis Krontiris. "User acceptance factors for anonymous credentials: An empirical investigation". In: *Proceedings of the 14th Annual Workshop on the Economics of Information Security (WEIS2015)*. 2015.

[110] Ashwin Machanavajjhala, Xi He, and Michael Hay. "Differential privacy in the wild: A tutorial on current practices & open challenges". In: *Proceedings of the VLDB Endowment* 9.13 (2016), pp. 1611–1614.

[111] Ahmad Sabouri. "On the user acceptance of privacy-preserving attribute-based credentials – a qualitative study". In: *International Workshop on Data Privacy Management (DPM 2016)*. Springer. 2016, pp. 130–145.

[112] Yang Liu and A.C. Simpson. "Privacy-preserving targeted mobile advertising: Formal models and analysis". In: *The 11th International Workshop on Data Privacy Management (DPM 2016)*. Springer. 2016, pp. 94–110.

[113] Chong Huang, Lalitha Sankar, and Anand D Sarwate. "Designing incentive schemes for privacy-sensitive users". In: *Journal of Privacy and Confidentiality* 7.1 (2016), p. 5.

[114] Imrul Kayes, Nicolas Kourtellis, Daniele Quercia, Adriana Iamnitchi, and Francesco Bonchi. "Cultures in community question answering". In: *Proceedings of the 26th ACM Conference on Hypertext & Social Media (HT 2015)*. ACM. 2015, pp. 175–184.

[115] M. C. Tschantz and J. M. Wing. "Formal methods for privacy". In: *Proceedings of the 2nd World Congress on Formal Methods (FM 2009)*. Ed. by A. Cavalcanti and D. Dams. Vol. 5850. Lecture Notes in Computer Science. Berlin/Heidelberg: Springer, 2009, pp. 1–15.

[116] Alan Abe and A.C. Simpson. "Formal Models for Privacy". In: *Proceedings of the 9th International Workshop on Privacy and Anonymity in the Information Society (PAIS 2016)*. Bordeaux, France, 2016.

[117] ISO/IEC. *ISO/IEC 13658: Information Technology—Z Formal Specification Notation—Syntax, Type System and Semantics*. ISO/IEC. 2002.

[118] D. Plagge and M. Leuschel. "Validating Z Specifications using the ProB Animator and Model Checker". In: *Proceedings of the 6th International Conference on Integrated Formal Methods (IFM 2007)*. Ed. by J. W. M. Davies and J. Gibbons. Vol. 4591. Lecture Notes in Computer Science. Springer, 2007, pp. 480–500.

[119] Google. *Google ad interest categories*. `https://support.google.com/ads/answer/2842480?hl=en&ref_topic=2971788`. [Last accessed March 2016].

[120] Hamed Haddadi, Heidi Howard, Amir Chaudhry, Jon Crowcroft, Anil Madhavapeddy, and Richard Mortier. "Personal Data: Thinking Inside the Box". In: *arXiv preprint arXiv:1501.04737* (2015).

[121] Ann Cavoukian. "Privacy by design: The 7 foundational principles". In: *Information and Privacy Commissioner of Ontario, Canada* (2009).

[122] Narseo Vallina-Rodriguez, Jay Shah, Alessandro Finamore, Yan Grunenberger, Konstantina Papagiannaki, Hamed Haddadi, and Jon Crowcroft. "Breaking for commercials: characterizing mobile advertising". In: *Proceedings of the 2012 Internet Measurement Conference (IMC 2012)*. ACM. 2012, pp. 343–356.

[123] L. L. C. SaurikIT. *Cydia Substrate*. `http://www.cydiasubstrate.com/`. [Last accessed September 2015].

[124] 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*. ACM. 2011, pp. 15–26.

[125] Jae-wook Jang, Jaesung Yun, Jiyoung Woo, and Huy Kang Kim. "Andro-profiler: Anti-malware system based on behavior profiling of mobile malware". In: *Proceedings of the companion publication of the 23rd international conference on World wide web companion*. International World Wide Web Conferences Steering Committee. 2014, pp. 737–738.

[126] Andrew Paverd, Andrew Martin, and Ian Brown. "Privacy-enhanced bi-directional communication in the Smart Grid using trusted computing". In: *Proceedings of the 5th IEEE International Conference on Smart Grid Communications (SmartGridComm 2014)*. IEEE. 2014, pp. 872–877.

[127] Wei-Hao Hwang, Yeong-Sheng Chen, and Tsang-Ming Jiang. "Personalized Internet Advertisement Recommendation Service Based on Keyword Similarity". In: *Proceedings of the 39th IEEE Annual Computer Software and Applications Conference (COMPSAC 2015)*. Vol. 1. IEEE. 2015, pp. 29–33.

[128] Anick Jesdanun. *Ad targeting based on ISP tracking now in doubt*. `http://www.mercurynews.com/business/ci_10337964`. [Last accessed October 2015]. 2008.

[129] Süleyman Barutçu. "Attitudes towards mobile marketing tools: A study of Turkish consumers". In: *Journal of Targeting, Measurement and Analysis for Marketing* 16.1 (2007), pp. 26–38.

[130] Nokia. *New Nokia research shows consumers ready for m-marketing via mobile handsets*. `http://company.nokia.com/en/news/press-releases/2002/01/30/new-nokia-research-shows-consumers-ready-for-m-marketing-via-mobile-handsets`. [Last accessed September 2015].

[131] Kai Wang, Shih-Hsiang Chen, and Hsin-Lu Chang. "The effects of forced ad exposure on the web". In: *Journal of Informatics & Electronics* 13.1 (2008), pp. 27–38.

[132] Yang Liu and A.C. Simpson. "AdSelector: A privacy-preserving advertisement-selection mechanism for mobile devices". In: *The Computer Journal* 60.8 (2017), pp. 1251–1270.

[133] Yahoo. *Yahoo ad interest manager*. `https://aim.yahoo.com/aim/us/en/optout/categories`. [Last accessed January 2016].

[134] Facebook. *How to target Facebook Adverts*. `https://www.facebook.com/business/a/online-sales/ad-targeting-details`. [Last accessed January 2016].

[135] Paul Pearce, Vacha Dave, Chris Grier, Kirill Levchenko, Saikat Guha, Damon McCoy, Vern Paxson, Stefan Savage, and Geoffrey M. Voelker. "Characterizing Large-Scale Click Fraud in ZeroAccess". In: *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security (CCS 2014)*. Scottsdale, Arizona, USA: ACM, New York, 2014, pp. 141–152.

[136] Vacha Dave, Saikat Guha, and Yin Zhang. "Measuring and Fingerprinting Click-spam in Ad Networks". In: *SIGCOMM Comput. Commun. Rev.* 42.4 (Aug. 2012), pp. 175–186.

[137] Allison Schiff. *Location inaccuracy is a bigger problem than fraud*. `http://adexchanger.com/mobile/location-inaccuracy-is-a-bigger-problem-than-fraud/`. [Last accessed January 2016]. 2015.

[138] Micro Speretta and Susan Gauch. "Personalized Search Based on User Search Histories". In: *Proceedings of the 2005 IEEE/WIC/ACM International Conference on Web Intelligence (WI 2005)*. WI '05. Compiègne, France: IEEE Computer Society, Washington DC, 2005, pp. 622–628.

[139] Google. *Google AdWords*. `http://www.google.co.uk/adwords/`. [Last accessed January 2016].

[140] G. Cho, J. Cho, Y. Song, and H. Kim. "An Empirical Study of Click Fraud in Mobile Advertising Networks". In: *Proceedings of the 10th International Conference on Availability, Reliability and Security (ARES 2015)*. Toulouse, France: IEEE Computer Society, Washington DC, Aug. 2015, pp. 382–388.

[141] Sumayah A. Alrwais, Alexandre Gerber, Christopher W. Dunn, Oliver Spatscheck, Minaxi Gupta, and Eric Osterweil. "Dissecting Ghost Clicks: Ad Fraud via Misdirected Human Clicks". In: *Proceedings of the 28th Annual Computer Security Applications Conference (ACSAC 2012)*. Orlando, Florida, USA: ACM, New York, 2012, pp. 21–30.

[142] Brendan Kitts, Jing Ying Zhang, Gang Wu, Wesley Brandi, Julien Beasley, Kieran Morrill, John Ettedgui, Sid Siddhartha, Hong Yuan, Feng Gao, et al. "Click Fraud Detection: Adversarial Pattern Recognition over 5 Years at Microsoft". In: *Real World Data Mining Applications*. Springer International Publishing, 2015, pp. 181–201.

[143] Google. *AdMobi behavioural policies*. `https://support.google.com/admob/answer/2753860`. [Last accessed January 2016].

[144] Google. *AdSense ad sizes available for mobile ads*. `https://support.google.com/adsense/answer/68727?hl=en-GB`. [Last accessed January 2016].

[145] ThinkWithGoogle. *Mobile App Marketing Insights: How Consumers Really Find and Use Your Apps*. `https://think.storage.googleapis.com/docs/mobile-app-marketing-insights.pdf`. [Last accessed January 2017]. 2015.

[146] MobileMarketing. *Average Brit has 27 apps installed, uses just six daily*. `http://mobilemarketingmagazine.com/average-brit-27-apps-installed-uses-just-six-daily/`. [Last accessed January 2017]. 2016.

[147] Usman Aleem, Hasan Cavusoglu, and Izak Benbasat. "An Empirical Investigation of the Antecedents and Consequences of Privacy Uncertainty in the Context of Mobile Apps". In: *Proceedings of the 16th Annual Workshop on the Economics of Information Security (WEIS2017)*. 2017.

[148] Badrul M Sarwar, Joseph A Konstan, Al Borchers, Jon Herlocker, Brad Miller, and John Riedl. "Using filtering agents to improve prediction quality in the grouplens research collaborative filtering system". In: *Proceedings of the 1998 ACM conference on Computer supported cooperative work*. ACM. 1998, pp. 345–354.

[149] David W McDonald and Mark S Ackerman. "Expertise recommender: A flexible recommendation system and architecture". In: *Proceedings of the 2000 ACM conference on Computer supported cooperative work*. ACM. 2000, pp. 231–240.

[150] Moon-Hee Park, Jin-Hyuk Hong, and Sung-Bae Cho. "Location-based recommendation system using bayesian user's preference model in mobile devices". In: *International Conference on Ubiquitous Intelligence and Computing*. Springer. 2007, pp. 1130–1139.

[151] Cisco. *Cisco Visual Networking Index: Global Mobile Data Traffic Forecast Update, 2016–2021 White Paper*. `https://www.cisco.com/c/en/us/solutions/collateral/service-provider/visual-networking-index-vni/mobile-white-paper-c11-520862.html`. [Last accessed November 2017]. 2017.

[152] Statista. *Internet of Things (IoT) connected devices installed base worldwide from 2015 to 2025*. `https://www.statista.com/statistics/471264/iot-number-of-connected-devices-worldwide/`. [Last accessed November 2017]. 2016.

[153] Juniper Research. *The Internet of Things: Consumer, Industrial & Public Services 2016—2021*. `https://www.juniperresearch.com/researchstore/iot-m2m/internet-of-things/consumer-industrial-public-services`. [Last accessed November 2017]. 2016.

[154] Federal Trade Commission. *Internet of Things — Privacy and Security in a Connected World*. https://www.ftc.gov/news-events/events-calendar/2013/11/internet-things-privacy-security-connected-world. [Last accessed November 2017]. 2013.

[155] George J Stigler. "An introduction to privacy in economics and politics". In: *The Journal of Legal Studies* 9.4 (1980), pp. 623–644.

[156] Alessandro Acquisti. "The Economics of Personal Data and the Economics of Privacy". In: *Economics* 11 (2010), pp. 24–2010.

[157] Kai-Lung Hui, Hock Hai Teo, and Sang-Yong Tom Lee. "The value of privacy assurance: An exploratory field experiment". In: *Mis Quarterly* (2007), pp. 19–33.

[158] Il-Horn Hann, Kai-Lung Hui, Sang-Yong Tom Lee, and Ivan PL Png. "Overcoming online information privacy concerns: An information-processing theory approach". In: *Journal of Management Information Systems* 24.2 (2007), pp. 13–42.

[159] Statista. *Number of smartphone users worldwide from 2014 to 2020*. https://www.statista.com/statistics/330695/number-of-smartphone-users-worldwide/. [Last accessed November 2017]. 2016.

[160] Statista. *Number of available applications in the Google Play Store from December 2009 to September 2017*. https://www.statista.com/statistics/266210/number-of-available-applications-in-the-google-play-store/. [Last accessed November 2017]. 2017.

[161] Statista. *Number of available apps in the iTunes App Store from 2008 to 2017*. https://www.statista.com/statistics/268251/number-of-apps-in-the-itunes-app-store-since-2008/. [Last accessed November 2017]. 2017.

[162] Zenith. *Advertising Expenditure Forecasts September 2017*. https://www.zenithmedia.com/insights-intelligence/global-intelligence/adspend-forecast/. [Last accessed November 2017]. 2017.

[163] Waleed A Alrodhan. "Privacy-Preserving Targeted Mobile Advertising Using Federated Identity Management Systems". In: *International Journal of Computer Science and Network Security (IJCSNS)* 17.10 (2017), p. 15.

[164] Franziska Roesner, Tohru Kohno, Alexander Moshchuk, Bryan Parno, Harry Jiannan Wang, and Crispin Cowan. "User-driven access control: Rethinking permission granting in modern operating systems". In: *Proceedings of the 2012 IEEE Symposium on Security and privacy (SP 2012)*. IEEE. San Francisco, CA, USA, 2012, pp. 224–238.

[165] Jonathan Jacky. "Model-Based Testing with Spec#". In: *Formal Methods and Software Engineering*. Springer, 2004, pp. 5–6.