Privacy-Preserving Targeted Mobile Advertising: Formal Models and Analysis

Yang Liu and Andrew Simpson

Department of Computer Science, University of Oxford
Wolfson Building, Parks Road, Oxford OX1 3QD
United Kingdom

Abstract. Targeted Mobile Advertising (TMA) has emerged as a significant driver of the Internet economy. TMA gives rise to interesting challenges: there is a need to balance privacy and utility; there is a need to guarantee that applications' access to resources is appropriate; and there is a need to ensure that the targeting of ads is effective. As many authors have argued, formal models are ideal vehicles for reasoning about privacy, as well as for reasoning about the relationship between privacy and utility. To this end, we describe how the formal notation Z has been used to develop formal models to underpin a prototype privacy-preserving TMA system. We give consideration to how formal models can help in underpinning the prototype system, in analysing privacy in the context of targeted mobile advertising, and in allowing users to specify control of their personal information.

1 Introduction

Targeted Mobile Advertising (TMA) is an important part of the Internet economy. By analysing personal information, organisations can deliver ads for specific goods and services that may be of interest to users. In [3], Beales indicates that the average quarterly pricing data for targeted advertising of 12 advertising networks was twice that for standard advertising in 2009. Further, in [17], Yan *et al.* suggest that the click-through rate of ads can be improved by, on average, 670% via the application of appropriate behavioural targeting strategies. However, TMA gives rise to privacy concerns: while users can take advantage of useful services, they are concerned about the misuse of their personal information and wish to not be 'tracked' [4]. The balance between concerns is, therefore, a delicate one.

There are two schools of thought with respect to trying to achieve this balance, with each school taking one 'side' or the other. On the one hand, researchers on the 'side' of corporations have tended to propose solutions that improve the collection of personal data and develop new analytical techniques to improve the accuracy of targeting (e.g. [2] and [18]). On the other hand, those on the users' 'side' tend to propose solutions that limit the ability of corporations to collect personal data (e.g. [5] and [8]). Our focus is a solution that tries to steer a middle path and that has the potential to be palatable to both users and corporations.

To this end, we have prototyped a system called *Privacy-Preserving Targeted Mobile Advertising* (PPTMA) [12], 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.

Such a system gives rise to a number of interesting challenges. First, there is a need to balance privacy and utility — and to do so in a way in which all parties can have confidence. Second, and relatedly, there is a need to ensure that all access to underlying resources by applications is appropriate. Finally, there is a need for a framework to support principled and effective selection of ads.

As argued by Tschantz and Wing [16], formal models have many roles to play in reasoning about privacy in a variety of contexts. As an example, in [1], 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." [16]

This contribution is in the spirit of that argument.

Formal models can be beneficial in many ways. In this paper, we give consideration to how formal models can help in underpinning our prototype system, PPTMA, in analysing privacy in the context of, 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 [10]. 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 [13], which supports both animation and model-checking.

We present our models in stages. 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. A further specification then describes our solution and helps to underpin our design. The final model is then 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.

2 Motivation and background

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. The targeting process is based on the user's data, which includes personal information, the

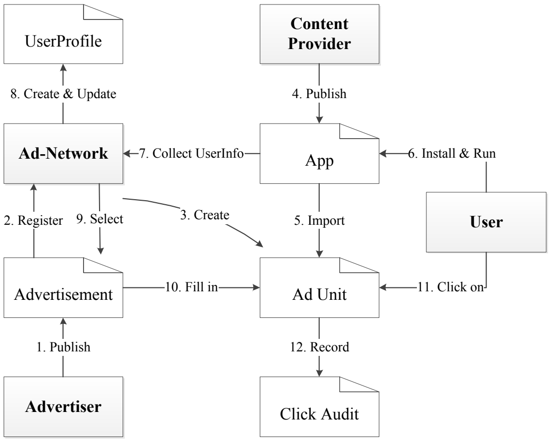

Fig. 1. A typical TMA workflow

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, the TMA system could present an ad of a sports store close to the user's current location.

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 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 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.

2.2 Motivation

The inherent tension between corporations and users is delicate: some researchers concern themselves with improving the performance of TMA systems (e.g. [2] and [18]), while others are concerned with the rights of users (e.g. [5] and [8]). Broadly, previous contributions in this area have sought to address the following questions:

1. *How to enhance the mobile advertising effectiveness for corporations?*
2. *How to preserve privacy for mobile users?*

Contributions that address problem 1 tend to disregard potential hostility from users; contributions that address problem 2 can lead to reduced benefits for all parties as a result of, for example, utilising fake user data.

In attempting to address these issues, some contributions also consider the following questions:

1R. *How to enhance the mobile advertising effectiveness for corporations, and reduce users' hostility?*

2R. *How to preserve privacy for mobile users, and enable them to take advantage of useful advertising services?*

A number of contributions (such as [6], [8], [9] and [15]) serve ads with a hybrid personalisation mechanism — pre-downloading ads from the ad server with a generalised context and selecting the most relevant one with respect to a fine-grained user profile maintained on the client. The hybrid approach allows corporations to deliver personalised ads without compromising mobile users' privacy. In addition, users can receive ads that are particularly useful; however, it is not easy for users to specify control over released personal information nor to understand the implications of the operations they perform in the ad-selection process. This gives rise to a further question:

3. *How to make the control of personal information easily specified by mobile users, and enable users to understand the effects of their decisions?*

These questions represent the primary motivation for this contribution — to present formal models that characterise a privacy-preserving TMA solution that has the potential to address questions 1R, 2R and 3. The models give confidence in our prototype 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.

3 A privacy-preserving solution

We now briefly introduce PPTMA [12]. 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 1.
3. **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 1.

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

We have implemented an initial prototype of PPTMA on the Android platform [12]. Some of the core challenges are handled as follows.

1. **API hooking.** Calling APIs is the main method for apps to collect users' personal information or execute permissions on the Android system. Therefore, we hook sensitive APIs at run-time to implement the functions of access control and monitoring of malicious apps.
2. **Feature library comparing.** Apps use ad-SDKs of ad-networks to collect user data and present ads on mobiles. To make use of ad-SDKs, content providers have to register their apps with ad-networks and import their libraries. By comparing feature codes of these libraries, we can deduce the particular ad-networks related to an app, the kinds of ad styles it contains, the potential behaviour it involves, etc.
3. **Ad-SDKs integration**. Feature library comparing enables the discovery of ad-SDKs contained by apps. For cooperative ad-networks, PPTMA imports the limited versions of their ad-SDKs to perform the basic functions for local TMA — pre-downloading ads lists by providing only limited anonymous user information and submitting view or click reports without specific user identifiers.

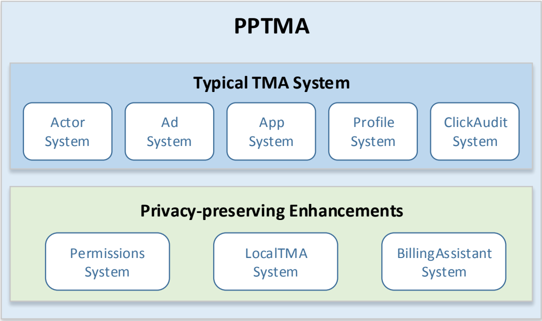

Fig. 2. The composition of the PPTMA system

4 Formal models

We now present brief overviews of the formal models of the PPTMA system and discuss how the formal models can help reason about issues of privacy. Figure 2 shows the composition of the overall system at a high level of abstraction. We start by considering the initial model of typical TMA systems to identify privacy-related behaviours.

4.1 A model of TMA

A typical TMA system is built up from many smaller components. In order to make our specification easier to grasp, 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 and constraints on state schemas.

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

$$\begin{array}{|l}
\hline
\mathit{ActorSystem} \\
\mathit{advertiser} : \mathit{AdvertiserId} \nrightarrow \mathit{Advertiser} \\
\mathit{adNetwork} : \mathit{AdNetworkId} \nrightarrow \mathit{AdNetwork} \\
\mathit{contentProvider} : \mathit{ContentProviderId} \nrightarrow \mathit{ContentProvider} \\
\mathit{user} : \mathit{UserId} \nrightarrow \mathit{User} \\
\hline
\end{array}$$

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.

$$\begin{array}{l}
\mathit{UserBasicInfo} \mathrel{\widehat{=}} [\, \mathit{gender} : \mathit{Gender};\ \mathit{age} : \mathit{Age}; \\
\qquad\qquad\qquad \mathit{location} : \mathit{Location};\ \mathit{language} : \mathit{Language} \,]
\end{array}$$

$$\begin{array}{l}\mathit{Ad} \mathrel{\widehat{=}} [\,\mathit{format} : \mathit{AdFormatId};\ \mathit{targetAudience} : \mathbb{P}\,\mathit{UserBasicInfo}; \\ \qquad\qquad \mathit{keyword} : \mathbb{P}\,\mathit{Keyword};\ \mathit{category} : \mathbb{P}\,\mathit{AdCategoryId}\,]\end{array}$$

$$\mathit{AdUnit} \mathrel{\widehat{=}} [\,\mathit{format} : \mathit{AdFormatId};\ \mathit{adNetwork} : \mathit{AdNetworkId}\,]$$

$$\begin{array}{|l}\hline \mathit{AdSystem} \\ \mathit{ad} : \mathit{AdId} \nrightarrow \mathit{Ad} \\ \mathit{adCategory} : \mathit{AdCategoryId} \nrightarrow \mathit{AdCategory} \\ \mathit{adUnit} : \mathit{AdUnitId} \nrightarrow \mathit{AdUnit} \\ \mathit{adFormat} : \mathit{AdFormatId} \nrightarrow \mathit{AdFormat} \\ \mathit{adInAdNetwork} : \mathit{AdNetworkId} \nrightarrow \mathbb{P}\,\mathit{AdId} \\ \hline \end{array}$$

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.

$$\begin{array}{|l}\hline \mathit{AppSystem} \\ \mathit{app} : \mathit{AppId} \nrightarrow \mathit{App} \\ \mathit{adUnitOfApp} : \mathit{AdUnitId} \nrightarrow \mathit{AppId} \\ \hline \end{array}$$

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

$$\begin{array}{l}\mathit{UserProfile} \mathrel{\widehat{=}} \\ \quad [\,\mathit{userBasicInfo} : \mathit{UserBasicInfo}; \\ \qquad \mathit{searchBrowseInfo} : \mathbb{P}\,\mathit{SearchBrowseInfo}; \\ \qquad\quad \mathit{selfMadeDocument} : \mathbb{P}\,\mathit{SelfMadeDocument}\,]\end{array}$$

$$\begin{array}{|l}\hline \mathit{ProfileSystem} \\ \mathit{userProfile} : \mathit{UserProfileId} \nrightarrow \mathit{UserProfile} \\ \mathit{userInterest} : \mathit{UserProfileId} \nrightarrow \mathbb{P}\,\mathit{AdCategoryId} \\ \mathit{profileOfUser} : \mathit{UserProfileId} \nrightarrow \mathit{UserId} \\ \mathit{profileInAdNetwork} : \mathit{UserProfileId} \nrightarrow \mathit{AdNetworkId} \\ \hline \end{array}$$

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.

$$\begin{array}{l}\mathit{ClickAudit} \mathrel{\widehat{=}} [\,\mathit{userId} : \mathit{UserId};\ \mathit{adId} : \mathit{AdId}; \\ \qquad\qquad\qquad \mathit{adUnitId} : \mathit{AdUnitId};\ \mathit{date} : \mathit{Date}\,]\end{array}$$

$$\begin{array}{|l}\hline \mathit{ClickAuditSystem} \\ \mathit{clickAudit} : \mathit{ClickAuditId} \nrightarrow \mathit{ClickAudit} \\ \mathit{clickAuditInAdNetwork} : \mathit{ClickAuditId} \nrightarrow \mathit{AdNetworkId} \\ \hline \end{array}$$

Combining these subsystems, we define a TMA system thus.

$$\begin{aligned}\mathit{System} \mathrel{\widehat{=}} [\,&\mathit{ActorSystem};\ \mathit{AdSystem};\\ &\mathit{AppSystem};\ \mathit{ProfileSystem};\ \mathit{ClickAuditSystem}\,]\end{aligned}$$

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, interests), which are stored in *ProfileSystem*, and users' behavioural data (e.g. browsing websites, 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, we can specify how much of a user's personal information is released to others.

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 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.

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 1.

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.

4.2 A model of PPTMA

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.

The permissions mechanism described (and implemented in our prototype) 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.

$$\begin{array}{|l}
\multicolumn{1}{l}{\mathit{PermissionsSystem}} \\
\hline
\mathit{permission} : \mathit{PermissionId} \nrightarrow \mathit{Permission} \\
\mathit{installedApp} : \mathit{UserId} \nrightarrow \mathbb{P}\,\mathit{AppId} \\
\mathit{permissionRequiredOfApp} : \mathit{AppId} \nrightarrow \mathbb{P}\,\mathit{PermissionId} \\
\mathit{permissionRequiredOfAdNetwork} : \mathit{AdNetworkId} \nrightarrow \mathbb{P}\,\mathit{PermissionId} \\
\mathit{permissionHeldOfInstalledApp} : (\mathit{UserId} \times \mathit{AppId}) \nrightarrow \mathbb{P}\,\mathit{PermissionId} \\
\hline
\end{array}$$

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, this 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*.

$$\begin{array}{|l}
\multicolumn{1}{l}{\mathit{LocalTMASystem}} \\
\hline
\mathit{customUserProfile} : \mathit{UserProfileId} \nrightarrow \mathit{UserProfile} \\
\mathit{localAds} : \mathit{UserId} \nrightarrow (\mathit{AdNetworkId} \nrightarrow \operatorname{seq} \mathit{AdId}) \\
\hline
\end{array}$$

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.

$$\begin{array}{|l}
\multicolumn{1}{l}{\mathit{BillingAssistantSystem}} \\
\hline
\mathit{obfuscatedClickAudit} : \mathit{ObfuscatedClickAuditId} \nrightarrow \mathit{ClickAudit} \\
\mathit{clickAuditMapping} : \mathit{ObfuscatedClickAuditId} \nrightarrow \mathit{ClickAuditId} \\
\hline
\end{array}$$

The PPTMA model is based on the three new subsystems, together with the model of the original TMA system.

$$\begin{aligned}
\mathit{PPTMA} \mathrel{\widehat{=}} [\, & \mathit{System};\ \mathit{PermissionsSystem}; \\
& \mathit{LocalTMASystem};\ \mathit{BillingAssistantSystem} \,]
\end{aligned}$$

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 local client. These operations can be implemented (in both stages) via custom algorithms.

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 locally 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.

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.

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.

$$\begin{aligned}\mathit{AudienceSegment} \mathrel{\widehat=} [\, & \mathit{userBasicInfo} : \mathit{UserBasicInfo}; \\ & \mathit{interestKeywords} : \mathbb{P}\, \mathit{Keyword} \,]\end{aligned}$$

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

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

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.

$$\begin{aligned}\mathit{UserBasicInfo}1 = \lparen\!\lvert\, & \mathit{gender} == \mathit{Male}, \mathit{age} == \mathit{actual}(25), \\ & \mathit{location} == \mathit{Oxford}, \mathit{language} == \mathit{English} \,\rvert\!\rparen\end{aligned}$$

We assume that profile ID of this user is *UserProfileId*1. The user is interested in *Basketball* (which we assume has the associated identifier *IdForBasketball*); therefore, in *ProfileSystem* the following predicate holds.

$$\mathit{IdForBasketball} \in \mathit{userInterest}(\mathit{UserProfileId}1)$$

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 *UserBasicInfo*1 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 following basic information:

   $$\begin{aligned} UserBasicInfo2 = \lblot\, & gender == DeclineToState, age == range(20, 30), \\ & location == UK, language == English \,\rblot \end{aligned}$$

   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.

   $$\begin{aligned} AudienceSegment1 = \lblot\, & userBasicInfo == UserBasicInfo1, \\ & interestKeywords == \{Basketball\} \,\rblot \\ AudienceSegment2 = \lblot\, & userBasicInfo == UserBasicInfo2, \\ & interestKeywords == \{TeamSports\} \,\rblot \end{aligned}$$

   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.

   $$\begin{array}{|l}
   \multicolumn{1}{l}{\_\_\, SelectAdsByAudienceSegment \,\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_} \\
   \Xi PPTMA \\
   as? : AudienceSegment \\
   anId? : AdNetworkId \\
   ads! : \mathbb{P}\, AdId \\
   \hline
   ads! = \{i : AdId \mid \\
   \qquad\qquad i \in (adInAdNetwork\ anId?) \land \\
   \qquad\qquad as?.userBasicInfo \in (ad\ i).targetAudience \land \\
   \qquad\qquad as?.interestKeywords \cap (ad\ i).keyword \neq \emptyset\} \\
   \hline
   \end{array}$$

   Here, the ad-network applies *AudienceSegment*2, which is abstracted from the coarse-grained user profile, as the input *as*?. Therefore, ads associates 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, publish 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.

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 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.

$$
\begin{array}{|l}
\multicolumn{1}{l}{\textit{SelectMostRelevantAds}} \\
\hline
\Xi \textit{PPTMA} \\
\textit{uId}? : \textit{UserId} \\
\textit{as}? : \textit{AudienceSegment} \\
\textit{auId}? : \textit{AdUnitId} \\
\textit{adsSet}! : \mathbb{P}\, \textit{AdId} \\
\textit{adsList}! : \mathrm{seq}\, \textit{AdId} \\
\hline
\textit{uId}? \in \mathrm{dom}\, \textit{localAds} \land \textit{auId}? \in \mathrm{dom}\, \textit{adUnit} \\
\textit{adsSet}! = \{\, i : \textit{AdId} \mid \\
\qquad\qquad (i \in \mathrm{ran}((\textit{localAds}\ \textit{uId}?)\ ((\textit{adUnit}\ \textit{auId}?).\textit{adNetwork})) \\
\qquad\qquad \land \\
\qquad\qquad \textit{as}?.\textit{userBasicInfo} \in (\textit{ad}\ i).\textit{targetAudience} \\
\qquad\qquad \land \\
\qquad\qquad \textit{as}?.\textit{interestKeywords} \cap (\textit{ad}\ i).\textit{keyword} \neq \emptyset \\
\qquad\qquad \land \\
\qquad\qquad (\textit{ad}\ i).\textit{format} = (\textit{adUnit}\ \textit{auId}?).\textit{format})\} \\
\textit{adsList}! = \\
\qquad ((\textit{localAds}\ \textit{uId}?)\ ((\textit{adUnit}\ \textit{auId}?).\textit{adNetwork})) \upharpoonright \textit{adsSet}! \\
\hline
\end{array}
$$

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.

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 4.2, 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.

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

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 [6, 7] are hardly clicked by humans, but 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 $BillingAssistantSystem$.

$$
\begin{array}{|l}
\multicolumn{1}{l}{\mathit{ClickFraudDetect}}\\
\Xi \mathit{PPTMA}\\
\mathit{ocId}? : \mathbb{P}\ \mathit{ObfuscatedClickAuditId}\\
\mathit{uId}! : \mathbb{P}\ \mathit{UserId}\\
\hline
\mathit{ocId}? \subseteq \operatorname{dom} \mathit{obfuscatedClickAudit}\\
\mathit{uId}! = \{u : \mathit{UserId} \mid\\
\qquad (\forall\, o? : \mathit{ObfuscatedClickAuditId} \bullet\\
\qquad\qquad u = (\mathit{clickAudit}\ (\mathit{clickAuditMapping}\ o?)).\mathit{userId})\}\\
\hline
\end{array}
$$

6 Analysis

We have used ProZ 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.

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 1 illustrates this. The analysis is based on the instance described in Section 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.

7 Conclusions

On the one hand, TMA provides significant financial benefits for advertisers. On the other hand, it gives rise to privacy concerns that users' personal information might be misused. Previous work in targeted advertising area (both on PCs and on mobile devices), such as Adnostic [15], Privad [6] and MobiAd [8], has typically tried to achieve the balance with a hybrid personalisation mechanism.

In this paper, 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 [16], 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.

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 1. Analysis on released personal information and related effects: example

Next steps will involve the development of a privacy-preserving ad-selection framework and related protocols, building on the existing prototype of [12]. The ad-selection framework allows ad-networks to apply their own algorithms in the pre-download and local selection processes; additional privacy-preserving protocols will be developed to ensure that no profile can be exposed in the communication between devices and ad-networks. We will also explore means of refining our access control model by leveraging work on user-driven access control (see, for example, [14]). Furthermore, we will continue to use our models to underpin model-based testing [11] as we further refine our prototype implementation.

References

1. Abe, A., Simpson, A.C.: Formal models for privacy. In: Proceedings of the 9th International Workshop on Privacy and Anonymity in the Information Society (PAIS 2016). Bordeaux, France (2016)
2. Ahn, H., Kim, K.J., Han, I.: Mobile advertisement recommender system using collaborative filtering: MAR-CF. In: Proceedings of the 2006 Conference of the Korea Society of Management Information Systems. pp. 709–715. The Korea Society of Management Information Systems (2006)

3. Beales, H.: The value of behavioral targeting. http://www.networkadvertising.org/pdfs/Beales_NAI_Study.pdf (2010), [Last accessed April 2015]
4. Farahat, A.: Privacy preserving frequency capping in Internet banner advertising. In: Proceedings of the 18th International Conference on World Wide Web (WWW 2009). pp. 1147–1148. ACM, Madrid, Spain (2009)
5. Goldfarb, A., Tucker, C.E.: Privacy regulation and online advertising. Management Science 57(1), 57–71 (2011)
6. Guha, S., Cheng, B., Francis, P.: Privad: Practical privacy in online advertising. In: Proceedings of the 8th USENIX Conference on Networked Systems Design and Implementation (NSDI 2011). pp. 169–182. Boston, MA, USA (2011)
7. Haddadi, H.: Fighting online click-fraud using bluff ads. ACM SIGCOMM Computer Communication Review 40(2), 21–25 (2010)
8. Haddadi, H., Hui, P., Brown, I.: MobiAd: Private and scalable mobile advertising. In: Proceedings of the 5th ACM International Workshop on Mobility in the Evolving Internet Architecture (MobiArch 2010). pp. 33–38. ACM, Chicago, IL, USA (2010)
9. Hardt, M., Nath, S.: Privacy-aware personalization for mobile advertising. In: Proceedings of the 2012 ACM conference on Computer and communications security (CCS 2012). pp. 662–673. ACM, Raleigh, NC, USA (2012)
10. ISO/IEC: ISO/IEC 13658: Information Technology — Z Formal Specification Notation — Syntax, Type System and Semantics. ISO/IEC (2002)
11. Jacky, J.: Model-based testing with spec#. In: Formal Methods and Software Engineering, pp. 5–6. Springer (2004)
12. Liu, Y., Simpson, A.C.: Privacy-preserving targeted mobile advertising: Requirements, design, and a prototype implementation. Software: Practice and Experience (2016), http://dx.doi.org/10.1002/spe.2403
13. Plagge, D., Leuschel, M.: Validating Z specifications using the ProB animator and model checker. In: Davies, J.W.M., Gibbons, J. (eds.) Proceedings of the 6th International Conference on Integrated Formal Methods (IFM 2007). Lecture Notes in Computer Science, vol. 4591, pp. 480–500. Springer (2007)
14. Roesner, F., Kohno, T., Moshchuk, A., Parno, B., Wang, H.J., Cowan, C.: User-driven access control: Rethinking permission granting in modern operating systems. In: Proceedings of the 2012 IEEE Symposium on Security and privacy (SP 2012). pp. 224–238. IEEE, San Francisco, CA, USA (2012)
15. Toubiana, V., Narayanan, A., Boneh, D., Nissenbaum, H., Barocas, S.: Adnostic: Privacy preserving targeted advertising. In: Proceedings of the 17th Annual Network and Distributed System Security Symposium (NDSS 2010). San Diego, CA, USA (2010), retrieved April 6, 2016 from https://www.isoc.org/isoc/conferences/ndss/10/pdf/05.pdf
16. Tschantz, M.C., Wing, J.M.: Formal methods for privacy. In: Cavalcanti, A., Dams, D. (eds.) Proceedings of the 2nd World Congress on Formal Methods (FM 2009). Lecture Notes in Computer Science, vol. 5850, pp. 1–15. Springer, Berlin/Heidelberg (2009)
17. Yan, J., Liu, N., Wang, G., Zhang, W., Jiang, Y., Chen, Z.: How much can behavioral targeting help online advertising? In: Proceedings of the 18th International Conference on World Wide Web (WWW 2009). pp. 261–270. ACM, Madrid, Spain (2009)
18. Yuan, S.T., Tsao, Y.W.: A recommendation mechanism for contextualized mobile advertising. Expert Systems with Applications 24(4), 399–414 (2003)