Privacy in targeted advertising: A survey

Imdad Ullah, Member, IEEE, Roksana Boreli, and Salil S. Kanhere, Senior Member, IEEE

Abstract—Targeted advertising has transformed the marketing landscape for a wide variety of businesses, by creating new opportunities for advertisers to reach prospective customers by delivering personalised ads, using an infrastructure of a number of intermediary entities and technologies. The advertising and analytics companies collect, aggregate, process and trade a vast amount of user’s personal data, which has prompted serious privacy concerns among both individuals and organisations. This article presents a detailed survey of the associated privacy risks and proposed solutions in a mobile environment. We outline details of the information flow between the advertising platform and ad/analytics networks, the profiling process, advertising sources and criteria, the measurement analysis of targeted advertising based on user’s interests and profiling context and the ads delivery process, for both in-app and in-browser targeted ads; we also include an overview of data sharing and tracking technologies. We discuss challenges in preserving user privacy that include threats related to private information extraction and exchange among various advertising entities, privacy threats from third-party tracking, re-identification of private information and associated privacy risks. Subsequently, we present various techniques for preserving user privacy and a comprehensive analysis of the proposals based on such techniques; we compare the proposals based on the underlying architectures, privacy mechanisms and deployment scenarios. Finally, we discuss the potential research challenges and open research issues.

Index Terms—Targeted advertising, Mobile advertising, Online behavioral advertising, Private information retrieval, Privacy, Information leakage, Privacy threats, Tracking, Private advertising systems, Billing, Cryptocurrency, Blockchain, RTB, Characterisation, Obfuscation, Differential privacy.

1 INTRODUCTION

Online advertising has become a prevalent marketing tool, commanding the majority of spending and taking over from the traditional broadcast advertising in newspapers, or television and radio. This is primarily due to the ability of online ad platforms to tailor or personalise ads, and thereby target specific customer segments. Targeted advertising is based on Big data analytics, where user’s personal information is collected and processed to enable segmenting users into groups based on interests, location, or personal attributes like age, gender, etc., with a varying size of the selected customer segment, down to the level of an individual.

The most significant platform from which personal data are collected and subsequently used for targeted ads is a mobile device, including mobile phones or tablets, due to it’s widespread and almost continuous use by a huge audience of potential ad recipients. A recent report [1] lists that 69% of user’s digital media time is actually spent on mobile phones only and consequently recommends tailoring targeted ads for mobile devices. Although the mobile users are still utilising browsers to access various online sites, applications (apps) are increasingly replacing the generic browser functionality. Currently, millions of mobile apps can be downloaded via various app marketplaces like the Google Play Store and the Apple App Store; it is projected that there will be more than 250 billion mobile app downloads by the end of 2021 [2].

Most mobile apps contain at least one ad library (including analytics’ libraries) [3] that enables targeted (or behavioural) mobile advertising to a wide range of audiences. The information about users and their online behaviour is collected through the ad library API calls [4], including information inference based on monitoring ads displayed during browsing sessions [5], [6]. The Advertising and Analytics (A&A) companies like Google Analytics and Flurry use this framework and are competing to increase their revenue by providing ad libraries that the apps developers use to serve ads. In the process of data monetisation, the advertising/analytics companies aggressively look for all possible ways to gather personal data from users, including purchasing users’ personal data from third parties.

The collection and use of personal data poses serious threats to privacy of users [7], [8], [9], [10], [11], [12], when websites or apps indicating sensitive information are used as the basis for profiling, e.g., a gaming app showing a gambling problem. Privacy concerns have been increasingly recognised by policy makers, with the introduction of anti-tracking laws, gradually making the use of some third-party tracking techniques used for interest-based targeting obsolete. E.g. Google has announced the Chrome’s ‘Cookie Apocalypse’, planning

1. Analytics is the systematic computational analysis of data or statistics for deeper understanding of consumer requirements. E.g. Google Analytics https://analytics.google.com, Flurry Analytics https://https://www.flurry.com/analytics/.
to phase out support for third-party cookies by 2022\textsuperscript{2}. Subsequently, instead of relying on third-party data, the A&A companies are increasingly using first-party data and shifting towards maintaining their own Data Management Platforms (DMPs) and Demand-Side Platforms (DSPs)\textsuperscript{3} to brand their own data and measure performance in a ‘cookie-less’ world. In a stronger push towards increased user’s privacy control over collection and use of their data, Apple\textsuperscript{4} has recently introduced the Identification for Advertisers (IDFA) opt-in overhaul in iOS 14.5, which will have significant impact on targeted ads and mobile ad/data attribution. This has created a very public feud with one of the largest social networks (and private data collection companies), Facebook [13], highlighting two different business approaches in regards to privacy and user targeting.

Overall, regardless of the technological and policy changes, protecting users’ personal data while having effective targeting is important to both the advertising networks and mobile users. Mobile users do want to view relevant (interest-based) ads, provided that their information is not exposed to the outside world including the advertising companies. Advertising networks can only be effective if they deliver the most relevant ads to users, to achieve better view/click through rates, while protecting the interactions between mobile users, advertisers and publishers/ad networks.

In this paper, we survey the threats and solutions related to privacy in mobile targeted advertising. We first present a survey of the existing literature on privacy risks, resulting from the information flow between the A&A companies, temporal tracking of users regarding both their activities and the outcomes of targeting them with personalised ads. We then describe, for both in-app (note that we interchangeably use ‘mobile’ and ‘in-app’) and in-browser targeted ads: the user profiling process, data collection and tracking mechanism, the ad delivery process and the process of ad characterisation. We outline the privacy threats posed by the A&A companies as a result of targeting; in particular, (to prove the privacy leakage) we demonstrate, using experimental evaluation, how private information is extracted and exchanged among various entities in an advertising system including third-party tracking and highlight the associated privacy risks. Subsequently, we provide an overview of privacy preserving techniques applicable to online advertising, including differential privacy, anonymisation, proxy-based solutions, k-anonymity i.e. generalisation and suppression, obfuscation, and crypto-based techniques such as Private Information Retrieval (PIR) and blockchain-based techniques. We also survey the proposed privacy preserving advertising systems and provide a comparative analysis of the proposals, based on the underlying architectures, the privacy techniques used and the deployment scenarios. Finally, we discuss the research challenges and open research issues.

This article is organised as follows. In Section 2, we introduce the mobile advertising ecosystem, its operation for ad delivery process, profiling process and characterisation of in-app and in-browser ads. Section 3 discusses the technical and in-depth understanding of ad network operations for targeted ads. Section 4 presents privacy threats and information leakage in online advertising systems. Section 5 presents a detailed comparative analysis of various privacy-preserving advertising systems. Various open research issues are outlined in Section 6. We conclude in Section 7.

2 **The Mobile Advertising Network**

The ad network ecosystem involves different entities which comprise of the advertisers, ad agencies and brokers, ad networks delivering ads, analytics companies, publishers and the end customers to whom ads are delivered [14]. For the case of large publishers, the ads may be served both by the publishers and the advertisers [15], consequently, the ad ecosystem includes a number of interactions between different parties.

2.1 The advertising ecosystem

A typical mobile ad ecosystem (both for in-app and in-browser ads) and the information flow among different parties is presented in Figure 1. A user has a number of apps installed on their mobile device, that are utilised with specific frequency. As demonstrated in [16], most mobile apps include analytics Software Development Kit (SDK) and as such both report their activity and send ad requests to the analytics and ad network. This network comprises the Aggregation server, analytics server, Billing server, and the Ads Placement Server (APS). Collected data, that relates to usage of mobile apps and the success of displayed ads, is used by the ads analytics server to develop user profiles (associated with specific mobile devices and corresponding users). A user profile comprises a number of interests, that indicates the use of related apps, e.g. sports, business, etc., constructed by e.g., Google Advertising network for Mobile (AdMob)\textsuperscript{5} and Flurry [17] (note that the latter is only visible to app developers). Targeted ads are served to mobile users according to their individual profiles. We note that other i.e., generic ads are also delivered [18]. The Billing server includes the functionality related to monetising Ad impressions (i.e. ads displayed to the user in specific apps) and Ad clicks (user action on selected ads); further discussion over ads billing is given in Section 2.5.

2. https://www.adviso.ca/en/blog/tech-en/cookie-apocalypse/
3. DMP is a unified and centralised technology platform used for collecting, organising, and activating large sets of data from disparate sources. DSP allows for advertisers to buy impressions across a number of different publisher sites, all targeted to specific users and based on key online behaviors and identifiers. See https://www.lotame.com/dmp-vs-dsp/ for detailed discussion over DMP and DSP.
4. https://junction.cj.com/article/button-weighs-in-what-does-apples-idfa-opt-in-overhaul-mean-for-affiliate
5. Google AdMob profile is accessible through the Google Settings system app on Android devices, accessible through Google Settings → Ads → Ads by Google → Ads Settings.
Develops apps
Mobile Device
App Developer
(1)
Creates mobile apps
Mobile app
(2)
Develops apps
Billing Server (BS)
(7)
Handles billing process
Analytics Server
(4)
Handles data analytics
Ad Placement Server (APS)
(6)
performs ad placement
Advertiser
(3)
Develops advertising
User Profiles
(8)
Maintains user profiles
Fig. 1: The mobile advertising ecosystem, including the information flow among different parties. Following functionalities: (1) Data collection and tracking, (2) Send tracking data to Aggregation server, (3) Forward usage info to Analytics server, (4) User profiling, (5) Send profiling info to APS, (6) Deliver targeted/generic ads, (7) Billing for apps developer, (8) Billing for Ad System, (9) Advertiser who wishes to advertise with Ad system.

2.2 User profiling

Advertising systems rely on user profiling and tracking to tailor ads to users with specific interests and to increase their advertising revenue. Following, we present the user profiling process, in particular, how the user profile is established, various criteria, and how it evolves over time.

2.2.1 Profile establishment

The advertising companies, e.g., Google, profile users based on the information they add to their Google account, data collected from other advertisers that partner with Google, and its estimation of user’s interests based on mobile apps and websites that agree to show Google ads. An example profile estimated by Google with various demographics (e.g. gender, age-ranks) and profiling interests (e.g. Autos & Vehicles) is shown in Figure 2. It is assumed that there is a mapping of the Apps profile $K_a$ (the apps installed on a user mobile device) to an Interests profile $I_g$ (such an example set of interests is shown in Figure 2) defined by advertising (e.g. Google) and analytics companies i.e. $K_a \rightarrow I_g$. This information is used by the analytics companies to individually characterise user’s interests across the advertising ecosystem.

This mapping includes the conversion of the apps categories $\Phi_j$ (where $j = 1, ..., \tau$ and $\tau$ is the number of different categories in a marketplace) to interest categories $\Psi_l$ (where $l = 1, ..., \epsilon$, $\epsilon$ is the number of interest categories defined by the analytics company). This mapping converts an app $a_{i,j} \in S_a$ to interests set $S_{i,j}^g$ after a specific level of activity $t_{est}$. The $t_{est}$ is the establishment threshold i.e. time an app should be used in order to establish profile’s interests. The result of this mapping is a set of interests, called Interests profile $I_g$. Google profile interests are grouped, hierarchically, under various interests categories, with specific interests.

In addition, the ads targeting is based on demographics so as to reach a specific set of potential customers that are likely to be within a specific age range, gender etc., Google presents a detailed set of various demographic targeting options for ads display, search campaigns etc. The demographics $D$ are usually grouped into different categories, with specific options such as age-ranges, e.g. ‘18-24’, ‘25-34’, ‘35-44’, ‘45-54’, ‘55-64’, ‘65 or more’, and gender e.g., ‘Male’, ‘Female’, ‘Rather not say’, and other options e.g. household income, parental status, location etc. The profiling is a result of interactions of user device with the AdMob SDK [8] that communicates with Google analytics for deriving user profiles. A complete set of ‘Web & App activities’ can be found under ‘My Google Activity’ [8], which helps Google make services more useful, such as, helping rediscover the things already searched for, read, and watched.

Figure 3 shows, a specific example of Google, various sources/platforms that Google use to collect data and target users with personalised ads. These include a wide range of different sources enabled with various tools, e.g., the ‘Web & Apps activities’ are extracted with the help of Andoird/iOS SDKs, their interactions with analytics servers within Google network, cookies, conversion tracking, web searches, user’s interactions with received

6. Google profile interests are listed in https://adsettings.google.com/authenticated?hl=en, displayed under the ‘How your ads are personalized’. Note that Google services can also be verified on Google Dashboard https://myaccount.google.com/dashboard?hl=en.
7. Demographic Targeting https://support.google.com/google-ads/answer/25603837?hl=en
8. https://myactivity.google.com/myactivity?otzr=1
9. https://support.google.com/google-ads/answer/6308
ads etc. Similarly, Google's connected home devices and services\textsuperscript{10} rely on data collected using cameras, microphones and other sensors to provide helpful features and services\textsuperscript{11}. Google Takeout\textsuperscript{12} can be used to export a copy of contents (up to several GBs of data) in user’s Google Account for backup or use it with a service outside of Google. Furthermore, this includes the data from a range Google products personalised for specific users that a user use, such as, email conversations (including ‘Spam’ and ‘Trash’ mails), contacts, calendar, browsing & location history, and photos.

2.2.2 Profile evolution

The profile is updated, and hence the ads targeting, each time variations in the users’ behavior are observed; such as for a mobile user using apps that would map to interests other than the existing set of interests. Let a user uses a new set of apps $S_u$, which has no overlap with the existing set of apps $S_a$ that has created $I_a$ i.e., $S_u \cap A = S_a$. $A$ is the set of apps in an app market. The newly added set of apps $S_u$ is converted to interests with $t_{ev}$ as evolution threshold i.e. the time required to evolve profile’s interests. Hence, the final Interests profile, $I_g'$, after the profile evolution process, is the combination of older interests derived during the profile establishment $I_a$ and during when the profile evolves $I_g''$.

2.2.3 Profile development process

In order for the Apps profile to establish an Interests profile, a minimum level of activity of the installed apps is required. Furthermore, in order to generate one or more interests, an app needs to have the AdMob SDK. We verified this by testing a total of 1200 apps selected from a subset of 12 categories, for a duration of 8 days, among which 1143 apps resulted the Interest profiles on all test phones indicating “Unknown” interests. We also note that the Apps profile deterministically derives an Interests profile i.e., a specific app constantly derives identical set of interests after certain level of activity. We further note that the level of activity of installed apps be within a minimum of 24hours period (using our extensive experiments; we note that this much time is required by Google analytics in order to determine ones’ interests), with a minimum of, from our experiments, 24/$n$ hours of activity of $n$ apps. For a sophisticated profiling, a user might want to install and use a good number of apps that would represent one’s interests. After the 24hours period, the profile becomes stable and further activity of the same apps does not result in any further changes. The mapping of Apps profile to Interests profile during the establishment and during the evolution process along with their corresponding stable states are shown in Figure 4.

Similarly, during the profile evolution process, the Interests profile starts changing by adding new interests; once apps other than the existing set of apps $S_a$ are utilised. However, instead of 24hours of period of evolving a profile, we observe that the evolution process adds additional interests in the following 72hours of period, after which the aggregated profile i.e. $I_g'$ becomes Stable. In order to verify the stability of the aggregated profile, we run these apps on 4th day; henceforth we observe no further changes. The mapping of Apps profile to Interests profile during the establishment and during the evolution process along with their corresponding Stable states are shown in Figure 4.

2.3 Targeted advertising

The mobile targeted advertising is a crucial factor in increasing revenue (a prediction shows the mobile ad market to grow to $408.58$ billion in 2026 \textsuperscript{19}) in a mobile app ecosystem that provides free services to the smartphone users. This is mainly due to users spend significantly more time on mobile apps than on the traditional web. Hence, it is important (note that targeted advertising is not only unique to the mobile ads but has also been used for in-browser to deliver ads based on user’s interests. The characterisation of targeted advertising, on the user’s side, is the in-depth analysis of the ad-delivery process so as to determine what information the mobile apps send to the ad network and how effectively they utilise this information for ads targeting. Furthermore, the characterisation of mobile targeted ads would expose the ad-delivery process and the ad networks can use the resultant analysis to enhance/redesign the ad delivery process, which helps in better view/click through rates.

It is crucial for the targeted advertising to understand as what information do apps (both free and paid mobile apps of various categories) send to the ad networks, in particular, how effectively this information is used to target users with interest-based ads? whether the ad networks differentiate among different types of users using apps from the same or different apps categories (i.e. according to Apps profile)? how much the ad networks differentiate mobile users with different profiles (i.e. according to Interests profile)? the effect over user profiling with the passage of time and with the use of apps from diverse apps categories (i.e. during profile evolution process)? the distribution of ads among users with different profiles? and the frequency of unique ads along with their ads serving distributions?

2.4 Ads selection algorithms

The accurate measurement of the targeted advertising is systematically related to the ad selection algorithm and is highly sensitive since it combines several fields of mathematics, statistics, analytics, and optimisation etc. Some of the ad selection algorithms show ad selection
A). Advertising and analytics platforms:
   1. Doubleclick Campaign manager
   2. Google Ads
   3. Google Analytics
   4. Youtube

B). Social media
C). Wearable devices
D). Google services
   e.g. Google Search

E). Web & Apps activities:
   1. Current search query
   2. Previous search activity/history
   3. Activity while a user is signed in to Google
   4. Previous interactions with ads
   5. Types of websites visited
   6. Types of mobile app activity on user device
   7. User activity on another device

F). Google account:
   1. Location
   2. Demographics, age-ranks, gender
   3. Time of day
   4. Info that user give to an advertiser, e.g.,
      signed up for a newsletter with email address

Fig. 3: Google’s data collection and tracking sources for targeting users with personalised ads (left) and tracking capabilities of analytics libraries enabled within mobile devices (right).

Fig. 4: Profile establishment & evolution processes. \( I_0 \) is the empty profile before apps utilisation. During the stable states, the Interest profiles \( I_g \) or \( I'_g \) remains the same and further activities of the same apps have no effect over the user profiles.

based on the user data pattern [20] and the program event analysis [21], however, the contextual and targeted advertising is treated in different way as they are related to the psyche of the users. Consequently, it has been observed that the activity of users and their demographics highly influences the ad selection along with the user clicks around an ad [22], [23]. As an example, a young female that is frequently browsing websites or using mobile apps related to the category of entertainment, would be more interested in receiving ads related to entertainment such as movies, musical instruments etc., consequently, it increases the click-through rates. Another work [24] builds a game-theoretic model for ad systems competing through targeted advertising and shows how it effects the consumers’ search behavior and purchasing decisions when there are multiple firms in the market. We note that the researchers utilise different ad selection and targeting algorithms based on machine learning and data mining techniques.

2.5 Ad billing

Billing is an important part of business models devised by any advertising system that is based on billing their customers for fine grained use of ad systems and their resources e.g. the advertisers set the payment settings and payment methods for monetising ad impressions and clicks. A number of studies show potential privacy threats posed by billing [25], [26], [27] i.e. a privacy-invasive architecture consists of service provides collecting usage information (such as particular interests of ads being shown and clicked) in order to apply appropriate tariff. Hence, among the important aims of private billing is to eliminate the leakage of private information and to minimise the cost of privacy across the billing period.

An example implementation of our private billing for ads, based on ZKP and Polynomial commitment (see detailed discussion over these techniques in Appendix B), is presented in [7], also shown in Figure 5. In this proposal, we presume that the following information is available to the client (software e.g. the AdMob SDK that is integrated in mobile apps for requesting ads and tracking user’s activity) for all ads in the database: the Ad index \( m \), Ad category \( \Phi_i \), price tags \( C^{prs}_T \) and \( C^{clk}_T \) respectively for ad presentations and ad clicks, and and the Advertiser ID \( ID_{Adv} \). This private billing mechanism consists of two parts: the work flow for retrieving ads (Step 1–3) and private billing (Step 4–13). In Step 2, the Ad server calculates the PIR response and sends it back to the client, following, the client decodes the PIR response (step 3) and forwards the retrieved ads to the mobile app.
Once the ad presentation (or ad click) process finishes then it undergoes the billing process. The client calculates the receipt locally, consisting of various components that are used to verify the following: (a) price tier for ad presented or ad clicks; (b) the \( ID_{Adv} \) (used for price deduction from advertiser, as shown in Step 11 of Figure 5); and (c) the application ID (helpful for price credit to App Developer i.e. Step 13). This billing mechanism is based on PS-PIR [27], proposed for e-commerce. We note that this billing mechanism is only applicable to single ad requests with no impact on privacy.

As opposed to above implementation, we suggested another proposal [28] for ad presentations and clicks with the use of mining Cryptocurrency (e.g. Bitcoin). The major aim for this proposal was for preserving user privacy, secure payment and for compatibility with the underlying AdBlock proposal [28] for mobile advertising system over Blockchain. Following notations are used in this proposal: price tags \( C_{Adv}^{pr} \) and \( C_{Adv}^{click} \) for ad presentation and click; various wallets i.e. App Developer’s \( wallet_{APP} \), Advertiser’s \( wallet_{AD} \), Billing server’s \( wallet_{BS} \); public-private key \( PK / PK’ \) and (Bitcoin) addresses, i.e. \( Add_{APP} / Add_{AD} / Add_{BS} \). It works as follows: The advertiser buys advertising airtime, it signs the message with the amount of Cryptocurrency with her private key \( PK’ \), adds Billing server’s address, requesting a transaction. Following, this request is bind with other transactions and broadcasted over the network for mining. Once the transaction completes, the Billing server receives its portion of Cryptocurrency in her wallet. In addition, the Miner initiates billing transaction for ads presentations or clicks respectively by encoding the \( C_{Adv}^{pr} \) and \( C_{Adv}^{click} \) price tags; this amount is then shared with \( wallet_{APP} \) and \( wallet_{AD} \) wallets.

### 3 Operations of Advertising System

Following, we discuss the technical aspects of the advertising systems e.g. the ad delivery process, ads traffic extraction and its characterisation, which eventually helps in understanding privacy issues in targeted advertising.

#### 3.1 Ad delivery process

We identify the workflow of a mobile app requesting a Google AdMob ad and the triggered actions resulting from e.g. a user click (we note that other advertising networks, such as Flurry, use different approaches/messages to request ads and to report ad clicks). Figure 6 describes some of the domains used by AdMob (Google ad servers and AdMob are shown separately for clarity, although both are acquired by Google). As shown, an ad is downloaded after the POST method is sent by mobile phone (Step 2) containing phone version, model, \( app \) running on phone etc. The ad contains the landing page (web address of an ad-URL) and JavaScript code that is executed where some of the static objects are downloaded (such as a PNG, Step 3). Two actions are performed after clicking an ad: a Conversion cookie\(^{13}\) is set inside phone (Step 4) and the web server associated with the ad is contacted. The landing page may contain other list of servers (mainly residing in Content Delivery Networks) where some of the static objects are downloaded and a complete HTML page is shown to the user (Step 5). The mobile apps developers agree on integrating ads in mobile apps and the ads are served according to various rules set by the ad networks, such as to fill up their advertising space, and/or obtaining profiling information for targeting. Additionally, the ads refreshment intervals, mechanisms used to deliver ads (push/pull techniques), the strategy adopted after ad is being clicked, and click-through rates etc. are also defined by the ad networks.

![Fig. 6: AdMob Ad Presentation Workflow [18.]](image)

In consequence, the ad networks are complex systems being highly diverse with several participants and adopting various mechanisms to deliver ads. Thus, in

---

\(^{13}\) Conversion tracking is specifically used by Google that is an action a customer takes on website that has value to the business, such as a purchase, a sign-up, or a view of a key page [29].
order to correctly identify and categorise ads and to server appropriate ads, it needs to investigate various ad delivery mechanisms and also cope with such diversity. This evaluation process needs identifying and collecting various ads delivery mechanisms through inspecting collected traffic traces captured from several apps executions, as shown in Figure 6. In addition, it needs to emphasise on ads distribution mechanisms used by ad networks from the apps’ perspective or users’ interests to devise the behaviour of ads pool served from ad networks and how they map to individual user’s interest profiles. Since the advertising system is a closed system, this process needs to indirectly evaluate the influence of different factors on ad delivery mechanisms, which is even more complicated in Real Time Bidding (RTB) scenarios and associated privacy risks.

3.2 Understanding ad network’s operation

The advertising networks provide an SDK for integrating ads inside the mobile apps while securing the low level implementation details. The ad networks provide regulation for embedding ads into the mobile apps, the ad delivery mechanism, the amount of times an ad is displayed on the user screen and how often an ad is presented to the user. The common type of ad is the flyer, which is shown to the user either at the top or at the bottom of device’s screen, or sometimes the entire screen is captured for the whole duration of the ad presentation. These flyers are composed of text, images and the JavaScript codes.

The ad presentation workflow of Google AdMob is shown in Figure 1 that shows the flow of information for an ad request by an app to AdMob along with the action triggered after the user clicks that particular ad. This figure shows the HTTP requests and the servers (i.e. Content Delivery Network (CDN) or ad servers) used by AdMob. Furthermore, several entities/services and a number of HTTP requests to interact with the ad servers and user agent can be observed in this figure.

3.3 Ad traffic analysis

3.3.1 Extracting ad traffic

Recall that the mobile ad network involves different entities to interact during the ad presentation and after an ad is being clicked to download the actual contents of the ad, as observed in Figures 1 and 6. Specifically, these entities are the products, the ad agencies attempting ad campaigns for the products, ad networks delivering ads, the publishers developing and publishing mobile apps, and the end customer to whom ads are delivered [14]. It is likely, when it comes to large publishers, that both the publishers and advertisers may have their own ad servers, in which case, some publishers may configure to put certain ads pool on the advertisers’ side and, at the same time, maintain their own ad servers [15]. The publishers, this way, can increase their revenue by means of providing redundant ad sources as if one ad network fails to deliver ads then they can try another ad network to continue providing services. In similar way, an end user may experience to be passed over several ad networks from publishers to the advertisers to access ads.

3.3.2 Ads traffic identification

The advertising system itself and its functionality are very diverse and complex to understand its operation [7], [30], hence in order to categorise the ad traffic, it needs to be able to incorporate such diversity. This can be performed by first capturing the traces from the apps that execute and download the ad traffic and then investigating the traffic characteristics. Characterising and inspecting the ad traffic can give information about the approaches used by multiple publishers, the various mechanisms used to deliver ads by the publishers, the use of different ad servers, and the ad networks themselves [28]. Similarly, it helps identify any analytics traffic used by the ad networks to target with relevant ads. Analysis of the traffic traces enables to parse and classify them as traffic related to i) ad networks, ii) the actual web traffic related to ad, iii) traffic related to CDN, iv) analytics traffic, v) tracking traffic, vi) ad auctions in RTB, and viii) statistical information about apps usage or developer’s statistics, and ix) traffic exchange during and after an ad click. As a consequence, a major challenge is to be able to derive comprehensive set of mechanisms to study the behaviours of ad delivery, classify the connection flows related to different ad networks, detecting any other possible traffic, and to classify them in various categories of ads.

3.3.3 Mobile vs. in-browser ads traffic analysis

We note that there are several differences in separately collecting and analysing the mobile and in-browser user’s ad/data traffic for the ad delivery mechanism in order to target users. Analysing the mobile ad traffic requires to be able to derive comprehensive set of rules to study the ad delivery behaviours (since several ad networks adopt their own formats for serving ads, as mentioned above), catalogue connection flows, and to classify ads categorisation. Furthermore, the ad delivery mechanisms are not publicly available, hence, analysing mobile targeted ads would be dealing with an inadequate information problem. Although in-browser ad delivery mechanism can be customised\textsuperscript{14} to receive ads which are tailored to a specific profiling interests [31], [32].

For the in-app ads delivery [7], [8], [33], [34], [35], an ad network may use different information to infer users’ interests, in particular, the installed applications together with the device identifier to profile users and to personalise ads pool to be delivered. Similarly, for in-browser ads, user profiling is performed by analytics

\textsuperscript{14} E.g. by modifying Google ads preferences: https://adssettings.google.com/authenticated?hl=en
companies [36] through different information such as browsing history, web searches etc., that is carried out using configured cookies and consequently target users with personalised ads. However, in in-app ad context, this information might be missing, or altogether not permitted by the OS, as the notion of user permissions may easily prevent the access to data out of the apps environment.

3.4 Characterisation of in-app advertisements

There is a limited research available on characterising the in-app (mobile) targeted ads. Prior research works have demonstrated the large extent to which apps are collecting user’s personal information [14], the potential implications of receiving ads to user’s privacy [6] and the increased utilisation of mobile device resources [15], [37]. In our previous study [18] (and in [38]), we observe that various information sent to the ad networks and the level of ads targeting are based on communicated information, similarly, we [9] investigate the installed apps for leaking targeted user data. To combat these issues, a number of privacy preserving [31], [32], [39] and resource efficient mobile advertising systems [15], [37] have been proposed. Works on the characterisation of mobile ads have primarily focused on measuring the efficiency of targeted advertising [22], to examine whether the targeted advertising based on the users’ behaviour leads to improvements in the click-through rates. However, thus far there have been limited insights about the extent to which targeting is effective in mobile advertising that will ultimately determine the magnitude of various issues such as bandwidth usage, including the loss of privacy.

We note that existing approaches on characterising targeted advertisements for in-browser [6], [22], [31], [32], [40], [41], [42], [43], [44], [45] cannot be directly applied to the evaluation of in-app ads due to the following reasons: First, the in-app targeting may be based on a number of factors that go beyond what is used for in-browser ads, including mobile apps installed on the device, the way they are utilised (e.g. heavy gamers may receive specific ads). Second, the classification of ads requires unifying of mobile market place(s) and traditional online environments, as the ads may relate both to merchant websites and to other apps that may be purchased and downloaded to the mobile devices. Third, the methodology for collecting information about in-app ads is different than for the in-browser ads, since the ad delivery process for in-app ads changes with every other ad network. Finally, apps come with pre-defined apps permissions to use certain resources, hence, allowing apps to filter part of the information to be provided to the ad network.

Figure 7 shows the lifecycle of characterising the ads traffic within the advertising system, both for in-app and in-browser targeted ads; various data scrapping elements and statistical measures are also shown on the right side of this figure.

Following we discuss few works on the characterisation of in-app and in-browser targeted ads.

3.4.1 In-app (mobile) ads

Few studies characterise various features of in-app ad traffic with the focus on targeted advertising. The MadScope [38] and [18] collects data from a number of apps, probes the ad network to characterise its targeting mechanism and reports the targeted advertising using profiles of specific interests and preferences. The authors in [37] analyse the ads harvested from 100+ nodes deployed at different geographic locations and 20 Android-based phones and calculated the feasibility of caching and pre-fetching of ads. The authors in [15] characterise the mobile ad traffic from numerous dimensions, such as, the overall traffic, the traffic frequency, and the traffic implications in terms of, using well-known techniques of pre-fetching and caching, energy and network signalling overhead caused by the system. This analysis is based on the data collected from a major European mobile carrier with more than three million subscribers. The [46] shows similar results based on the traces collected from more than 1,700 iPhone and Windows Phone users.

The authors in [47] show that apps from the same category share similar data patterns, such as geographic coverage, access time, set of users etc., and follow unique temporal patterns e.g. entertainment apps are used more frequently during the night time. The [48] performs a comparative study of the data traffic generated by smartphones and traditional internet in a campus network. Another work [49], studies the cost overhead in terms of the traffic generated by smartphones that is classified into two types of overheads i.e. the portion of the traffic related to the advertisements and the analytics traffic i.e. traffic transmitted to the third-party servers for the purpose of collecting data that can be used to analyse users’ behaviour etc. Several other works, [50], [51], [52], study profiling the energy consumed by smartphone apps.

3.4.2 In-browser ads

There are a number of works on characterising in-browser ads with the focus on issues associated with the user privacy [42], [44]. In [6], the authors present classifications of different trackers such as cross-site, in-site, cookie sharing, social media trackers, and demonstrate the dominance of tracking for leaking user’s privacy, by reverse engineering user’s profiles. They further propose a browser extension that helps to protect user’s privacy. Prior research works show the extent to which consumers are effectively tracked by third parties and across multiple apps [53], mobile devices leaking Personally Identifiable Information (PII) [54], [55] and apps accessing user’s private and sensitive information through well defined APIs [56]. Another study [57] reveals by using differential correlation technique in order to identify various tracking information used for targeted ads. Similarly, [58] investigates the ad fraud that generates spurious revenue affecting the ad agencies. In
addition, other studies, such as [59] describes challenges in measuring online ad systems and [45] provides a general understanding of characteristics and changing aspects of advertising and targeting mechanisms used by various entities in an ad ecosystem.

4 PRIVACY IN MOBILE ADVERTISING: CHALLENGES

Privacy can be defined as “the ability of an individual or group to seclude themselves or information about themselves, and thereby express themselves selectively”\(^\text{15}\). In addition, the Personally Identifiable Information (PII) is the “information that can be used to distinguish or trace an individual’s identity”\(^\text{16}\), which if compromised or disclosed without authorisation, may result in harm, embarrassment, inconvenience, or unfairness to an individual. Recall that the profiling and targeted advertising expose potentially sensitive and damaging information about users, also demonstrated in [60], [61], [62]. There is a growing user awareness of privacy and a number of privacy initiatives, e.g., Apple’s enabling of ad blockers in iOS\(^\text{17}\) is representative of a move towards giving users greater control over the display of ads, although applicable only to browser based rather than to mobile targeted ads, however, this would greatly affect Google’s services, since Google’s services are now based on Web & App activity\(^\text{18}\).

Hence, the purpose of targeted advertising is to be able to protect user’s privacy and effectively serve relevant ads to appropriate users, in particular, to enable private profiling and targeted ads without revealing user interests to the advertising companies or third party ad/ tracking companies. Furthermore, an private billing process to update the advertising network about the ads retrieved/ clicked in a privacy preserving manner.

4.1 Privacy attacks

There are various kinds of privacy attacks, we mainly focus on three main categories of privacy attacks. Note that in all these scenarios, the user is not opposed to profiling in general and is willing to receive services e.g., targeted ads, on selected topics of interests, but does not wish for specific parts of their profile (attributes), based on the usage of apps (s)he considers private, to be known to the analytics network or any other party, or to be used for personalised services.

4.1.1 Unintended privacy loss

In this case, users voluntary provide personal information, e.g. to OSNs, or users authorize third-party services to access personal information, e.g. third-party library tracking in mobile apps, however users may not be aware how the information is used and what are the potential privacy risks.

4.1.2 Privacy leakage via cross-linking or de-anonymisation

The user profile is (legitimately) derived by the analytics network (e.g. [7], [8], [9] focused on Google AdMob
and Flurry) by cross-linking private information or via de-anonymisation. In the former case, the analytics services aggregate user data from sources that supposedly come as a results of users (willingly) previously shared their data with various data owners for providing them personalised services. In the later case, the data owners release anonymised personal information or data sources that sell data to advertisers or data anonymised data freely available on various websites\(^\text{19}\). The anonymised data is used to leak privacy when attackers disclose the identity of the data owner by cross-linking to external data sources i.e. using background knowledge [9].

### 4.1.3 Privacy leakage via statistical inference
The statistical inference i.e., an indirect attack over user privacy, that involves a third party profile users based on their behavior to provide personalised services e.g., the advertising systems e.g., Google or Flurry monitor the ad traffic [9], [18] sent to mobile devices and infers the user profile based on their targeted ads. The profiling attributes are sensitive to the users and are considered as private information e.g. political view, religious, sexual orientation, etc.

### 4.2 Ad traffic analysis for evaluating privacy leakage
Several works investigate the mobile targeted ads traffic primarily for the purpose of privacy and security concerns. The AdRisk [3], an automated tool, analyse 100 ad libraries and studies the potential security and privacy leakages of these libraries. The ad libraries involve the resource permissions, permission probing and JavaScript linkages, and dynamic code loading. Parallel to this work, [63] examines various privacy vulnerabilities in the popular Android-based ad libraries. They categorise the permissions required by ad libraries into optional, required, or un-acknowledged and investigate privacy concerns such as how user’s data is sent in ad requests. The authors in [64] analyse the privacy policy for collecting in-app data by apps and study various information collected by the analytics libraries integrated in mobile apps.

Other works [65], [66] study the risks due to the lack of separate working mechanisms between Android apps and ad libraries and propose methods for splitting their functionality. The authors in [14] monitor the flow of data between the ad services and 250K Android apps and demonstrate that currently proposed privacy protecting mechanisms are not effective, since app developers and ad companies do not show any concern about user’s privacy. They propose a market-aware privacy-enabling framework with the intentions of achieving symmetry between developer’s revenue and user’s privacy. Another work [67] carried out a longitudinal study in the behaviour of Android ad libraries, of 114K free apps, concerning the permissions allocated to various ad libraries over time. The authors found that over several years, the use of most of the permissions has increased over time raising privacy and security concerns.

There has been several other works, exploring the web advertisements in different ways i.e. form the monetary perspective [22], [68], from the perspective of privacy of information of users [69], from privacy information leakage and to propose methods to protect user data [70], [71], and the E-Commerce [72]. In similar way, a detailed analysis of the web ad networks from the perspective information communicated on network level, the network layer servers, and from the point of the content domains involved in such a system are investigated [73].

### 4.3 Inference of private information
In recent years, several works [74], [75], [76], [77], [78], [79], [80], [81], [82] have shown that it is possible to infer undisclosed private information of subscribers of online services such as age, gender, relationship status, etc. from their generated contents. The authors in [78] analysed the contents of 71K blogs at blogger.com and were able to accurately infer the gender and age of the bloggers. The authors were able to make their inferences by identifying certain unique features pertaining to an individual’s writing style such as parts-of-speech, function words, hyper-links and content such as simple content words and the special classes of words taken from the handcrafted LIWC (Linguistic Inquiry and Word Count) [83] categories.

Another study [74] has shown that the age demographics of Facebook users (both using apps and browsers) can be predicted by analysing the language used in status update messages. Similar inferences have been made for IMDB users based on their movie reviews [79]. Another work [81] predicts age, gender, religion, and political views of users from the queries using models trained from Facebook’s ‘Like’ feature. In [76], the authors analysed client-side browsing history of 250K users and were able to infer various personal attributes including age, gender, race, education and income. Furthermore, a number of studies [84], [85], [86] have demonstrated that sensitive attributes of user populations in online social networks can be inferred based on their social links, group memberships and the privacy policy settings of their friends [87].

### 4.4 User information extraction
We experimentally evaluate [9] how to extract user profiles from mobile analytics services based on the device identifier of the target; this method was demonstrated using both Google analytics and Flurry in the Android environment. Here the user profile, i.e. set of information collected or inferred by the analytics services, consists of personally identifiable information such as, unique device ID, demographics, user interests inferred from the app usage etc.
An crucial technique to extract user profiles from the analytics services (we mainly target Google and Flurry analytics services) is to first impersonate the victim’s identity; then Case 1 Google analytics: to fetch user profiles from a spoofed device, where the private user profile is simply shown by the Google service as an ads preference setting or Case 2 Flurry analytics: to inject the target’s identity into a controlled analytics app, which impacts those changes in the Flurry audience analysis report using which the adversary is able to extract user profile. Following, we first describe how to obtain and spoof a device’s identity, subsequently, the user profile extraction for both cases of Google and Flurry is presented in detail.

4.4.1 Information extraction via user profiles from Google

Android system allows users to view and manage their in-app ads preferences20, e.g. to opt-out or to update/delete interests. This feature retrieves user profile from Google server which is identified by the advertising ID. As a consequence of the device identity spoofing, an adversary is able to access the victim’s profile on a spoofed device.

We note that there are at least two possible ways to that an adversary can capture victims Android ID. First, an adversary can intercept the network communication, in order to capture the usage reporting messages sent by third-party tracking APIs, extract the device identifier, and to further use it for ongoing communication with the analytics services. Note that it is very easy to monitor IDs of thousands of users in a public hotspots e.g. airport, hotel etc. Similarly, in a confined area, an adversary (e.g. an employer or a colleague) targeting a particular individual can even associate the collected device ID to their target (e.g. employees or another colleague). During this privacy attack, we note that Google analytics library prevents leakage of device identity by hashing the Android IDs; however it cannot stop other ad libraries to transmit such information in plain text (which can be easily be mapped to Google’s hashed device ID).

An alternative way, although may be more challenging in practice, is to obtain the target’s device identifier from any application (controlled by the adversary) that logs and exports the device’s identity information.

4.4.2 Information extraction via user profiles from Flurry

We note that extracting user profiles from Flurry is more challenging since Flurry does not directly allow users to view or edit user’s Interests profiles. In fact, except the initial consent on the access of device resources, many smartphone users may not be aware of the Flurry’s tracking activity.

Figure 8 shows the basic operations of our profile extraction technique within the mobile advertising ecosystem. To compromise a user’s private profile, an adversary spoofs the target device, identified by deviceIDa, using another Android device or an emulator. Following, the adversary uses a bespoke app with a (legitimate) appIdx, installed on the spoofed device, to trigger a usage report message to Flurry. Accordingly, the analytics service is manipulated into believing that deviceIda is using a new application tracked by the system. Consequently, all user related private information is made accessible to the adversary through audience analysis report of appIdx in Flurry system.

An adversary can easily extract the corresponding statistics and link them to (legitimate) user once the audience report from Flurry targets a unique user. In addition, the adversary will be able to track and access all subsequent changes to the user profile at a later time. In our presented technique, since we do impersonate a particular target’s device ID, we can easily associate the target to a ‘blank’ Flurry-monitored application.

Alternatively, an adversary can derive an individual profile from an aggregated audience analysis report by monitoring report differences before and after a target ID has been spoofed (and as such has been added to the audience pool). Specifically, the adversary has to take a snapshot of the audience analysis report Pt at time t, impersonates a target’s identity within his controlled Flurry-tracked application, and then takes another snapshot of the audience analysis report at Pt+1. The target’s profile is obtained by extracting the difference between Pt and Pt+1, i.e. ∆(Pt, Pt+1). However in practice, Flurry service updates profile attributes on a weekly basis which means it will take up to a week to extract a full profile per user.

Finally, the segment feature provided by Flurry, the app audience is further split by applying filters according to e.g. gender, age group and/or developer defined parameter values. This feature allows an adversary to isolate and extract user profiles in a more efficient way. For instance, a possible segment filter can be ‘only show users who have Android ID value of x’ which results in the audience profile containing only one particular
user. The effectiveness of the attack are validated in two steps: 1. We first validate that user’s profile is the basis for ads targeting, by showing that specific profiles will consistently receive highly similar ads and conversely, that a difference in a user’s profile will result in a mobile receiving dissimilar ads. 2. We then perform the ad influence attack, i.e. we perturb selected profiles and demonstrate that the modified profiles indeed receive in-app ads in accordance with the profile modifications.

4.5 Third-party privacy threats

The third-party A&A libraries have been examined in a number of works, such as [3], [15], [16], [63], [88], which contribute to the understanding of mobile tracking and collecting and disseminating personal information in current mobile networks. The information stored and generated by smartphones, such as call logs, emails, contact list, and GPS locations, is potentially highly sensitive and private to the users. Following, we discuss various means through which users’ privacy is exposed.

4.5.1 Third-party tracking

Majority of privacy concerns of smartphone users are because of inadequate access control of resources within the smartphones e.g. Apple iOS and Android, employ fine-grained permission mechanisms to determine the resources that could be accessed by each application. However, smartphone applications rely on users to allow access to these permissions, where users are taking risks by permitting applications with malicious intentions to gain access to confidential data on smartphones [89]. Similarly, privacy threats from collecting individual’s online data (i.e. direct and inferred leakage), have been examined extensively in literature, e.g. [10], [90], including third party ad tracking and visiting [91], [92].

Prior research works show the extent to which consumers are effectively tracked by a number of third parties and across multiple apps [53], mobile devices leaking PII [54], [55], apps accessing user’s private and sensitive information through well defined APIs [56], inference attacks based on monitoring ads [9] and other data platform such as eXelate21, BlueKai22, and AddThis23 that collect, enrich and resell cookies.

The authors in [93] conducted a user survey and showed that minor number of users pay attention to granting access to permissions during installation and actually understand these permissions. Their results show that 42% of participants were unaware of the existing permission mechanism, only 17% of participant paid attention to permissions during apps installation while only 3% of participants fully understood meaning of permissions accessing particular resources. The authors in [3] evaluate potential privacy and security risks of information leakage in mobile advertisement by the embedded libraries in mobile applications. They studied 100,000 Android apps and identified 100 representative libraries in 52.1% of apps. Their results show that the existing ad libraries collect private information that may be used for legitimate targeting purposes (i.e., the user location) while other data is harder to justify, such as the users call logs, phone number, browser bookmarks, or even the list of apps installed on the phone. Additionally, they identify some libraries that use unsafe mechanisms to directly fetch and run code from the Internet, which also leads to serious security risks. A number of works [94], [95], [96], identify the security risks on Android system by disassembling the applications and tracking the flow of various methods defined within various programmed classes.

There are several works to protect privacy by assisting users to manage permissions and resource access. The authors in [97] propose to check the manifest files of installed mobile apps against the permission assignment policy and blocking those that request certain potentially unsafe permissions. The MockDroid [98] track the resource access and rewrites privacy-sensitive API calls to block information communicated outside the mobile phones. Similarly, the AppFence [99] further improves this approach by adding taint-tracking, hence, allowing more refined permission policies.

4.5.2 Re-identification of sensitive information

Re-identification involves service personalisation based on pervasive spatial and temporal user information that have already been collected e.g. locations that users have already visited. The users are profiled and later on provided with additional offers based on their interests, such as, recommending on places to visit, or people to connect to. There have been a number of research works to identify users based on re-identification technique. For instance, the authors in [100] analyse U.S. Census data and show that on average, every 20 individuals from the dataset share same home or work locations while 5% of people in dataset can be uniquely identified by home-work location pairs. Another related work [101] uniquely identifies US mobile phone users using generalisation technique by generalising the top N homework location pairs. They use location information to derive quasi-identifiers for re-identification of users. Similarly, a number of research works e.g. [102], [103], [104], raise privacy issues in publishing sensitive information and focus on theoretical analysis of obfuscation algorithms to protect user privacy.

4.6 Quantifying privacy algorithms

Quantifying privacy is an important and challenging task as it is important to evaluate the level of privacy.

21. https://microsites.nielsen.com/daas-partners/partner/exelate/
22. https://www.oracle.com/corporate/acquisitions/bluekai/
23. https://www.addthis.com/
5 Privacy in Mobile Ads: Solutions

The direct and indirect (i.e., inferred) leakages of individuals’ information have raised privacy concerns. A number of research works propose private profiling (and advertising) systems [32], [39], [116], [117], [118], [119]. These systems do not reveal either the users’ activities or the user’s interest profiles to the ad network. Various mechanisms are used to accomplish these goals: Adnostic [32], Privad [117] and Re-priv [116] focus on targeting users based on their browsing activities, and are implemented as browser extensions running the profiling algorithms locally (in the user’s browser). MobiAd [39] proposes a distributed approach, specifically aimed at mobile networks. The use of differential privacy is advocated in Practical Distributed Differential Privacy (PDDP) [118] and SplitX [119], where differentially private queries are conducted over distributed user data. All these works protect the full user profile and advocate the use of novel mechanisms that necessitate the re-design of some parts or all of the current advertising systems, although some (e.g., Adnostic) can operate in parallel with the existing systems. In addition, the works based on the use of noisy techniques like differential privacy, to obfuscate user’s preferences may result in a lower accuracy of targeted ads (and correspondingly lower revenues), compared to the use of standard targeting mechanisms.

Figure 9 shows the lifecycle of proposal for privacy-preserving mobile/web advertising systems; specifically starting from data collection for evaluating privacy/security risks, baseline model and proposed business model for preserving user’s privacy, finally model evaluation and its comparison with the baseline model. Various data scrapping elements, statistical measures and privacy preserving techniques are also shown in this figure.

An important thing in the development of private advertising system is that the consumers’ trust in privacy of mobile advertising is positively related to their willingness to accept mobile advertising [120], [121]. The AdChoices25 program (a self-regulation program implemented by the American ad industry), states that consumer could opt-out of targeted advertising via online choices to control ads from other networks. However, another study [122] examines that the opt-out users cause 52% less revenue (and hence presents less relevant ads and lower click through rates) than those users who allow targeted advertising. In addition, the authors noted that these ad impressions were only requested by 0.23% of American consumers.

5.1 Private ad ecosystems

There are a number of generic privacy preserving solutions proposed to address the negative impact of ads targeting. Anonymity solutions for web browsing include the use of Tor [123], or disabling the use of cookies [124]. These accomplish the goal of preventing user tracking, however, they also prevent any user (profile based) service personalisation, that may actually be a desirable feature for many users despite their privacy concerns.

Research proposals to enable privacy preserving advertising have been more focused on web browsing, as the dominant advertising media e.g., [32], [33], [117], [119], [125], propose to use locally derived user profiles. In particular, Privad [117] and Adnostic [32] use the approach of downloading a wide range of ads from the ad network and locally (in the browser or on the mobile device) selecting ads that match the user’s profile. On the other hand, there are a smaller number of works address privacy for mobile advertising, with representative works e.g., [7], [8], [28], [34], [39], [126], [127], suggest the app-based user profiling, stored locally on mobile device. The [7] is based on various mechanisms of PIR and it complements the existing advertising system and is conceptually closest to [126], which uses Oblivious RAM (ORAM) to perform Private Information Retrieval (PIR) on a secure coprocessor hardware. However, unlike our

25. https://optout.aboutads.info/?c=2&lang=EN
solution it relies on specific (secure) hardware to enable PIR, which may limit its applicability in a general setting.

5.2 Data masking, anonymisation, obfuscation and randomisation

There are several privacy protection techniques, such as techniques based on anonymisation e.g. encrypting or removing PII, proxy-based solutions, k-anonymity i.e. generalisation and suppression, obfuscation (making the message confusing, willfully ambiguous, or harder to understand), mechanisms based on differential privacy i.e. maximising the accuracy of queries from statistical databases while minimising the chances of identifying its records, crypto-based techniques such as private information retrieval (PIR) and blockchain-based solutions. Following we present various privacy-preserving advertising systems based on these different techniques.

5.2.1 Anonymisation

The simplest and most straightforward way to anonymise data includes masking or removing data fields (attributes) that comprise PII. These include direct identifiers like names and addresses, and quasi-identifiers (QIDs) such as gender and zip code, or an IP address; the later can be used to uniquely identify individuals. It is assumed that the remainder of the information is not identifying and therefore not a threat to privacy (although it contains information about individuals, e.g. their interests, shopping patterns, etc.). A second approach is to generalise QIDs, e.g., by grouping them into a higher hierarchical category (e.g., locations into post codes); this can also be accomplished according to specified generalisation rules. Anonymisation mechanisms that deal with selected QIDs according to pre-determined rules include k-anonymity [128] and it’s variants like l-diversity [107] and t-closeness [108]. These, in their simplest form, k-anonymity (detailed discussion over k-anonymity is given in Appendix C), modifies (generalise) individual user records so that they can be grouped into identical (and therefore indistinguishable) groups of k, or additionally apply more complex rules (l-diversity and t-closeness).

A number of proposals advocate the use of locally (either in the browser of the mobile device) derived user profiles, where user’s interests are generalised and/or partially removed (according to user’s privacy preferences), before being forwarded to the server or an intermediary that selected the appropriate ads to be forwarded to the clients. In the context of targeted advertising, the removal of direct identifiers includes user IDs (replacing them with temporary IDs) or mechanisms to hide used network address (e.g., using TOR [123]). However, if only the most obvious anonymisation is applied without introducing additional (profiling and targeting oriented) features, the ad networks ecosystem would be effectively disabled. Therefore, we only mention representative solutions from this category and concentrate on the privacy-preserving mechanisms that enable targeted ads.

The privacy requirements are also, in a number of prior works, considered in parallel with achieving bandwidth efficiency for ad delivery, by using caching mechanisms [37], [39], [117]. Furthermore, such techniques have been demonstrated to be vulnerable to composition attacks [129], and can be reversed (with individual users identified) when auxiliary information is available (e.g. from online social networks or other publicly available
In Adnostic [32], each time a webpage (containing ads) is visited by the user; the client software receives a set of generic ads, randomly chosen by the broker. The most appropriate ads are then selected locally, by the client, for presentation to the user; this is based on the locally stored user profile. We have categorised this work as a generalisation mechanism as the served ads are generic (non-personalised), although it could arguably be considered under the randomisation techniques. We note that in [32] the user’s privacy (visited pages or ad clicks) is not protected from the broker.

In Privad [31], [117], a local, (detailed) user profile is generated by the Privad client and then generalised before sending to the ads broker in the process of requesting (broadly) relevant ads. All communication with the broker is done through the dealer, which effectively performs the functions of an anonymising proxy; the additional protection is delivered by encrypting all traffic, this protecting user’s privacy from the dealer. The proposed system also includes monitoring of the client software to detect whether any information is sent to the broker using, e.g., a covert channel. Similarly, in MobiAd [39], the authors propose a combination of peer-to-peer mechanisms that aggregates information from users and only presents the aggregate (generalised activity) to the ad provider, for both ad impressions and clicks. Caching is utilised to improve efficiency and Delay tolerant networking for forwarding the information to the ad network. Similarly, another work [132] proposes combining of users interests via an ad-hoc network, before sending them to the ad server.

Additionally, some system proposals [133] advocate the use of anonymisation techniques (l-diversity) in the targeting stage, where the ads are distributed to users, while utilising alternative mechanisms for profiling, learning and statistics gathering.

5.2.2 Obfuscation

Obfuscation is the process of obscuring the intended meaning of the data or communication by making the message difficult to understand.

In the scenario of an advertising system, recall that the user privacy is mainly breached for their context i.e. specific use of mobile apps from an app category, and their profiling interests along with the ads targeting based on these interests. Hence, an important focus in implementing such mechanisms is to obfuscate specific profiling attributes that are selected as private (i.e. the attributes that the analytics companies may use for interest-based advertisements) and the categories of installed apps. For instance, the user may not wish the categories of gaming or porn to be included in their profile, as these would reflect heavy use of corresponding (gaming and porn) apps. The obfuscation scenarios can be based on similar (obfuscating) apps or similar profiling attributes or interests customised to user’s profile [8] or randomly chosen apps/interests from non-private categories. An important factor is to take into consideration the extra (communication, battery, processing, usage airtime) overhead while implementing obfuscation mechanisms, following, it needs present jointly optimised framework that is cost effective and preserves user privacy for profiling, temporal apps usage behavioral patterns and interest-based ads targeting.

A recent work [134] carries out a large scale investigation of obfuscation use where authors analyse 1.7 million free Android apps from Google Play Store to detect various obfuscation techniques, finding that only 24.92% of apps are obfuscated by the developer. There are several obfuscation mechanisms for protecting private information, such as the obfuscation method presented in [135] that evaluates different classifiers and obfuscation methods including greedy, sampled and random choices of obfuscating items. They evaluate the impact of obfuscation, assuming prior knowledge of the classifiers used for the inference attacks, on the utility of recommendations in a movie recommender system. A practical approach to achieving privacy [136], which is based on the theoretical framework presented in [137], is to distort the view of the data before making it publicly available while guaranteeing the utility of the data. Similarly, [138] proposes an algorithm for publishing partial data that is safe against the malicious attacks where an adversary can do the inference attacks using association rule in publicly published data.

Another work, ‘ProfileGuard’ [34] and its extension [8] propose an app-based profile obfuscation mechanism with the objective of eliminating the dominance of private interest categories (i.e. the prevailing private interest categories present in a user profile). The authors provide insights to Google AdMob profiling rules, such as showing how individual apps map to user’s interests within their profile in a deterministic way and that AdMob requires a certain level of activity to build a stable user profile. These works use a wide-range of experimental evaluation of Android apps and suggest various obfuscation mechanisms e.g. similarity with user’s existing apps, bespoke (customised to profile obfuscation) and bespoke++ (resource-aware) strategies. Furthermore, the authors also implement a POC ‘ProfileGuard’ app to demonstrate the feasibility of an automated obfuscation mechanism.

Following, we provide an overview of prior work in both randomisation (generic noisy techniques) and differentially private mechanisms.

5.2.3 Randomisation

In the randomisation methods, noise is added to distort user’s data. Noise can either be added to data values (e.g., movie ratings or location GPS coordinates), or, more applicable to profiling and user targeting, noise is in the form of new data (e.g., additional websites that the user would not have visited normally are generated by a browser extension [139]), added in order to mask the true vales of the records (browsing history). We note that...
[139] protects the privacy of user’s browsing interests but does not allow (privacy preserving) profiling or selection of appropriate targeted ads.

The idea behind noise addition is that specific information about user’s activities can no longer be recovered, while the aggregate data still contains sufficient statistical accuracy so that it can be useful for analysis (e.g., of trends). A large body of research work focuses on generic noisy techniques e.g., [140] proposed the approach of adding random values to data, generated independently of the data itself, from a known e.g., the uniform distribution. Subsequent publications (e.g., [141]) improve the initial technique, however other research work [142] has identified the shortcomings of this approach, where the added noise may be removed by data analysis and the original data (values) recovered.

A novel noisy technique for privacy preserving personalisation of web searches was also recently proposed [143]. In this work, the authors use ‘Bloom’ cookies that comprise a noisy version of the locally derived profile. This version is generated by using Bloom filters [144], an efficient data structure; they evaluate the privacy versus personalisation trade-off.

5.3 Differential privacy

The concept of differential privacy [26] was introduced in [145], a mathematical definition for the privacy loss associated with any released data or transcript drawn from a database. Two datasets D1 and D2 differ in at most one element given that one dataset is the subset of the other with larger database contains only one additional row e.g., D2 can be obtained from D1 by adding or removing a single user. Hence, a randomised function K gives differential privacy for the two data sets D1 and D2 as: P_r [K (D1) ∈ S] ≤ exp (ε) × P_r [K (D2) ∈ S]. We refer readers to [146] for deeper understanding of differential privacy and its algorithms.

Differential privacy is vastly used in the literature for anonymisation e.g., a recent initiative to address the privacy concerns by recommending usage of differential privacy [147] to illustrate some of the short-comings of direct contact-tracing systems. Google has recently published a Google COVID-19 Community Mobility Reports [27] to help public health authorities understand the mobility trends over time across different categories of places, such as retail, recreation, groceries etc., in response to imposed policies aimed at combating COVID-19 pandemic. The authors in [148] use differential privacy to publish statistical information of two-dimensional location data to ensure location privacy. Other works, such as [149], [150], partition data dimensions to minimise the amount of noise, and in order to achieve higher privacy accuracy, by using differential privacy in response to the given set of queries.

Differential privacy [151] work has, in recent years, resulted in a number of system works that advocate the practicality of this, previously predominantly theoretical research field. The authors in [118] propose a system for differentially private statistical queries by a data aggregator, over distributed users data. A proxy (assumed to be honest-but-curious) is placed between the analyst (aggregator) and the clients and secure communications including authentication and traffic confidentiality are accomplished using TLS [152]. The authors also use a cryptography solution to provide additional privacy guarantees. The SplitX system [119] also provides differential privacy guarantees and relies on intermediate nodes, which forward and process the messages between the client that locally stores their (own) data and the data aggregator. Further examples include works proposing the use of distributed differential privacy [153] and [154].

5.4 Cryptographic mechanisms

A number of different cryptographic mechanisms have been proposed in the context of profiling and targeted advertising or, more broadly, search engines and recommender systems. These include: Private Information Retrieval (PIR), Homomorphic encryption, Multi-party Computing (MPC), Blockchain based solutions.

5.4.1 Private Information Retrieval (PIR)

Private Information retrieval (PIR) [110], [111], [115], [155], [156], [157], is the ability to query a database successfully without the database server discovering which record(s) of the database was retrieved or the user was interested in. Detailed discussion over various PIR mechanisms along with their comparison is given in Appendix A.

The ObliviAd proposal [126] uses a PIR solution based on bespoke hardware (secure coprocessor), which enables on-the-fly retrieval of ads. The authors propose the use of Oblivious RAM (ORAM) model, where the processor is a “black box”, with all internal operations, storage and processor state being unobservable externally. ORAM storage data structure comprises of entries that include a combination of keyword and a corresponding ad (multiple ads result in multiple entries). The accounting and billing are secured via the use of using electronic tokens (and mixing [158], [159]). More generally, a system that enables private e-commerce using PIR was investigated in [27], with tiered pricing with record level granularity supported via the use of the proposed Priced Symmetric PIR (PS-PIR) scheme. Multiple sellers and distributed accounting and billing are also supported by the system.

Additionally, cryptographic solutions can be used to provide part of the system functionality. They are commonly used in conjunction with obfuscation, e.g., in [153], [154] or generalisation [32].
5.4.2 Zero Knowledge Proof (ZKP) and Mixing

Zero knowledge proofs [160], [161], [162], [163] and mixing [164] are commonly used as components of the privacy solutions. ZKP is a cryptographic commitment scheme by which one party (the prover) can prove to another party (the verifier) that they know a value $x$, without conveying any information apart from the fact that they know the value $x$. An example of Mixing, called mixnet [158], based on cryptography and permutation, was introduced to achieve anonymity in network communication. It creates a hard-to-trace communication by using a chain of proxy servers, called mixes, which takes messages from multiple senders, shuffle, and send them back in random order to the destination, hence, breaking the link between source and destination and making it harder for eavesdroppers to trace end-to-end communications. A number of robust, threshold mix networks have appeared in the literature [159], [165], [166], [167], [168], [169], [170].

Chen et al. [118] uses cryptographic mechanism to combine client-provided data (modified in accordance with differential privacy). They utilise a probabilistic Goldwasser-Micali cryptosystem [171]. In their subsequent work [119], the authors use an XOR-based cryptographic mechanism to provide both anonymity and unlinkability to analysis (queries) of differentially private data distributed on user’s devices (clients). A cryptography technique, mixing [158], [159] is also commonly used as part of anonymisation [126], [172], where mix servers are used as intermediaries that permute (and re-encrypt) the input.

5.4.3 Homomorphic encryption

Homomorphic encryption [173] is a form of encryption that allows specific types of computations to be carried out on ciphertext, without decrypting it first, and generates an encrypted result that, when decrypted, matches the result of operations performed on the plaintext.

Adnostic [32] uses a combination of homomorphic encryption and zero-knowledge proof mechanisms to enable accounting and billing in the advertising system in a (for the user) privacy preserving way. Effectively, the user is protected as neither the publisher (website that includes the ads) nor the advertisers (that own the ads) have knowledge about which users viewed specific ads. The authors in [153] also combine differential privacy with a homomorphic cryptosystem, to achieve privacy in a more generic setting of private data aggregation of distributed data. Similarly, Shi et al. [154] also use a version of homomorphic techniques to enable private computing of sums based on distributed time-series data by a non-trusted aggregator.

The authors in [174] presents privacy-preserving recommendations using partially homomorphic encryption (PHE) along with secure multi-party computation protocols. Specifically, user’s private data encrypted via PHE, this way the recommender cannot use their original data while still being able to generate private recommendation, is uploaded to the recommender system; following the recommender runs a cryptographic protocol offline with a third party to generate personalised recommendations. This proposal also achieves good performance by lowering the processing and communication overheads by borrowing high cryptographic computations from third-party systems. Similarly, [175] proposes a recommendation system based on the ElGamal cryptosystem (i.e. a kind of PHE), where all users actively collaborate with recommender server privately generate recommendations for a target user. Another work [176] relies on Boneh-Goh-Nissim (BGN) homomorphic cryptosystem that adopts an additional isolated recommender server that assists users in decrypting ciphertexts whenever necessary, hence, actively interact with both recommendation and additional servers.

5.4.4 Multi-Party Computing (MPC)

MPC [177] is a set of cryptographic methods that allow private computing (of selected mathematical functions) on data from multiple, distributed, parties, without exposing any of the input data. The formal guarantees provided by MPC relate to both data confidentiality and the correctness of the computed result.

A web-based advertising system was first proposed by Juels [172], where they use multi-party information-theoretic (threshold) PIR in an honest-but-curious multi-server architecture. Central to their system is the choice of a Negotiant function, that is used by the advertiser to select ads, starting from a user’s profile - the authors describe both a semi-private and a fully private information-theoretic (threshold) PIR in an honest-but-curious multi-server architecture. They evaluate the benefits of both alternatives in regards to security, computational cost and communication overheads. In addition, in one of our previous works [7], our motivation for using information-theoretic (threshold) PIR for mobile private advertising system, rather than other solutions, e.g., Oblivious Transfer [178], [179], is the lower communication and computation overheads of such schemes.

5.5 Blockchain-based advertising systems

Blockchain is a fault-tolerant distributed system based on a distributed ledger of transactions, shared across the participating entities, and provides auditable transitions [180], where the transactions are verified by participating entities within operating network. A blockchain is unalterable i.e. once recorded, the data in any block cannot be changed without altering of all the subsequent blocks; hence, it may be considered secure by design with high Byzantine fault tolerance e.g., one quarter of the participating nodes can be faulty but the overall system continues to operate normally.

Among the participating entities in a blockchain-based network; the Miner is a special node responsible for generating transactions, adding them to the pool of
pending transactions and organizing into a block once the size of transactions reaches a specific block size. The process of adding a new block to the Blockchain is referred to as mining and follows a consensus algorithm, such as Proof of Work (POW) [181] and Proof of Stake (POS) [182], which ensures the security of Blockchain against malicious (Miner) users. The participating entities use the Public-Private Key pair that is used to achieve the anonymity [183]. Among various salient features of Blockchain, i.e. irreversible, audible, updated near real-time, chronological and timestamp, which, in addition, disregards the need of a central controlling authority; thus making it a perfect choice for restricting communication between the mobile apps and the analytics/ad companies and keeping individual’s privacy.

Blockchain [184] has numerous applications and has been widely used, e.g. IoT [185], Bid Data [186], Healthcare [187], Banking and finance [188] etc. Blockchain has become a new foundation for decentralised business models, hence in the environment of advertising platform, made it a perfect choice for restricting communication between mobile apps (which is potentially a big source of private data leakage) and the ad/analytics companies and keeping individual’s privacy.

To our knowledge, we note that there are very limited works available for Blockchain-based mobile targeted ads in the literature e.g. the [35] presents a decentralised targeted mobile coupon delivery scheme based on Blockchain. The authors in this work match the behavioral profiles that satisfy the criteria for targeting profile, defined by the vendor, with relevant advertisements. However, we note that this framework does not include all the components of an advertising system including user profiles construction, detailed structure of various Blockchain-based transactions and operations, or other entities such as Miner and the billing process. Our recent work, AdBlock [28], presents a detailed framework (in addition to Android-based POC implementation i.e. a Bespoke Miner) for privacy preserving user profiling, privately requesting ads, the billing mechanisms for presented and clicked ads, mechanism for uploading ads to the cloud, various types of transactions to enable advertising operations in Blockchain-based network, and methods for access policy for accessing various resources, such as accessing ads, storing mobile user profiles etc. This framework is parented in Figure 10. We further experimentally evaluate its applicability by implementing various critical components: evaluating user profiles, implementing access policies, encryption and decryption of user profiles. We observe that the processing delays with various operations evaluate to an acceptable amount of processing time as that of the currently implemented ad systems, also verified in [7].

Summary of various privacy preserving approaches, in terms of architecture, mechanism, deployment and app domain, for both in-browser and mobile advertising systems is given in Table 1.

Table 2 provides a hypothetical comparison of various privacy protection mechanisms for various important
TABLE 1: Summary of the in-browser and in-app advertising systems.

| Ref | Architecture | Mechanism | Deployment | Domain |
|-----|--------------|-----------|------------|--------|
| privad [17] | 3rd-party anonymising proxy | Crypto | Browser add-on | Web |
| Adnostic [32] | Complements to existing sys | Crypto billing | Firebox extension | Mobile |
| PASTE [152] | Untrusted third party | Fourier Perturbation Algo | Browser add-on | Client/Server sides |
| [389] | Cookie management | User preference | Standalone | |
| [390] | Anonymising proxy | Differential privacy | HTTP header | Browser side |
| BNTI [19] | Delay Tolerant Network | Encryption | Mobile phone | |
| MobiAd [39] | Complements existing sys | Crypto-based | Client/Server sides | Mobile |
| ObliviAd [126] | | | |
| [127] | | | |
| SPiDX [19] | Context prediction | Crypto-based | Blockchain | |
| ProfileGuard [8], [34] | Profile Oblfuscation | | | |
| [55] | | | | |
| AdBlock [28] | | | | |
| [7] | Autonomous system | Crypto-based | Standalone | |

TABLE 2: Comparison of various privacy protection mechanisms for various important parameters applicable in an advertising system.

| Parameters | Differential Privacy | Obscuration | Cryptographic mechanisms | Randomisation | Blockchain solutions | Anonymisation |
|------------|---------------------|-------------|--------------------------|---------------|----------------------|--------------|
| Apps usage behavioral privacy | No guarantee | Guaranteed | Yes | No guarantee | No guarantee | No guarantee |
| Profiling privacy | Low (Low to high) | No guarantee | Yes | No guarantee | No guarantee | No guarantee |
| Protect privacy exposure from targeted ads | Yes | Yes | Yes | No | Yes | No |
| Cost of achieving user privacy | Low | High | Low | High | Low | Low |
| Targeted ads | Yes (Lower) | Limited to no relevant ads (adjustable) | Limited to no relevant ads (adjustable) | Yes | Limited to no relevant ads (adjustable) | Yes |
| Tradestyles privacy and targeted ads | No | Yes | Yes | No | Yes | No |
| Impact over timing for targeted ads | Yes | Yes | Yes | No | Yes | No |

It can be observed that the **Obfuscation**-based mechanisms can guarantee user’s ‘apps usage behavior privacy’ (as evident in [8], [34]) at the expense of installing and running a number of mobile apps, similarly, the ‘cost’ of achieving user privacy with **Blockchain**-based solution is quite high due to its operational complexity [28], [35]. An important parameter is ‘impact over targeted ads’ as a result of running a number of mobile apps with various techniques e.g., **Crypto-based** techniques (such as PIR), **Blockchain** and **Anonymisation** techniques will have no impact over targeted ads, alternatively, the **Differential privacy**, **Obfuscation** and **Randomisation** will have a significant impact over targeted ads and can be adjusted according to the user’s needs i.e., ‘low-relevant vs. high-relevant interest-based ads’, as is also evident in [8], [9]; note that these latter set of techniques will also have impact over billing since the advertisers’ ads are shown to “irrelevant” users, hence, they (advertisers) pay for airtime that is used by non-targeted audiences. Similarly, an important parameter is the ‘trade-off between privacy and targeted ads’, which can only be achieved using the **Obfuscation** and the **Randomisation** techniques. Furthermore, another parameter is to protect user privacy in terms of serving targeted ads i.e., an ‘indirect privacy attack to expose user privacy’, which cannot be exposed when **Crypto-based** techniques are used since the delivered ads are also protected, as shown in [7].

### 5.7 The economic aspects of privacy

Research works also investigate the notion of compensating users for their privacy loss, rather than imposing limits on the collection and use of personal information.

Ghosh and Roth [192] studied a market for private data, using **differential privacy** as a measure of the privacy loss. The authors in [193] introduce transactional privacy, which enables the users to sell (or lease) selected personal information via an auction system. On a related topic of content personalisation and in-browser privacy, in RePriv [116] the authors propose a system that fits into the concept of a marketplace for private information. Their system enables controlling the level of shared (local) user profile information with the advertising networks, or, more broadly, with any online entity that aims to personalise content.

### 6 Open Research Issues

In this section, we present various future research directions that require further attention from the research community i.e. diffusion of user data in Real Time Bidding (RTB) scenarios and associated privacy risks, the complicated operations of advertising system, the user-driven private mobile advertising systems and its private billing mechanism.
6.1 Diffusion of user tracking data

A recent shift in the online advertising has enabled by the advertising ecosystem to move from ad networks towards ad exchanges, where the advertisers bid on impressions being sold in RTB auctions. As a result, the A&A companies closely collaborate for exchanging user data and facilitate bidding on ad impressions and clicks [194], [195]. In addition, the RTB cause A&A companies to perform additional tasks of working with publishers to help manage their relationship for ad exchange (in addition to user’s tracking data) and to optimise the ad placement (i.e. targeted ads) and bidding on advertiser’s behalf. This has made the online advertising operations and the advertising ecosystems themselves extremely complex.

Hence, it is important for the A&A companies to model (in order to accurately capture the relationship between publisher and A&A companies) and evaluate the impact of RTB on the diffusion of user tracking (sensitive) data. This further requires assessing the advertising impact on the user’s contexts and profiling interests, which is extremely important for its applicability and scalability in the advertising scenarios. This will also help the A&A companies and publisher to effectively predict the tracker domain and to estimate their advertising revenue. Furthermore, to ensure the privacy of user data since the data is collected and disseminated in a distributed fashion i.e. users affiliated to different analytics and advertising platforms and shared their data across diverse publishers. This also necessitates a distributed platform for the efficient management and sharing of distributed data among various A&A platforms and publishers. In particular, the RTB has demanded to develop efficient methods for distributed and private data management.

6.2 Complex operations of advertising system

The complexity of online advertising poses various challenges to user privacy, processing-intensive activities, interactions with various entities (such as CDN, analytics servers, etc.) and their tracking capabilities. In order to reduce the complexity of the advertising systems, we envision few more areas of research: devising processing-sensitive frameworks, limiting the direction-redirection of requests among A&A entities, unveil user data exchange processes within the ad platform, identifying new privacy threats and devising new protection mechanisms. Unveiling user data exchange will expose the extent to which the intermediate entities prone to adversarial attacks. Hence, it requires a better knowledge of adversary, which will contribute to develop protection mechanisms for various kinds of privacy threats, such as, interest-based attacks, direct privacy attacks. Note that this will further require comparative analysis of basic and new proposals for the trade-off achieved between privacy and computing overheads of processing user’s ad retrieval requests/responses, communication bandwidth consumption and battery consumption.

6.3 Private user-driven mobile advertising systems

An enhanced user-driven private advertising platform is required where the user interest (vis-à-vis their privacy) and advertising system’s business interests may vary, in addition, the assessment of user information as an inherent economic value will help to study the tradeoff between such values and user privacy within the advertising system. This will require the proposal for complex machine learning techniques to enhance ads targeting (since previous works found that majority of received ads were not tailored to intended user profiles [18], [38], which will ultimately help advertising systems to increase their revenues and enhance user experience in receiving relevant ads. Likewise, introducing novel privacy preserving mechanisms, a very basic step would be to combine various proposals, as described in Section 5, which will introduce more robust and useful privacy solutions for various purposes: enhanced user targeting, invasive tracking behaviors, better adapting privacy enhancing technologies, better adapt the changing economic aspects and ethics in ads targeting. Another research direction would be to extend the analysis of privacy protection mechanisms to other different players, such as, advertisers, ad exchange, publishers with the aim to analyse and evaluate privacy policies and protection mechanisms that are claimed by these parties. This would help various entities in the advertising system to identify the flaws and further improve their working environment.

Another research direction would be to create smarter privacy protection tools on the user side i.e. to create such tools as an essential component of mobile/browser-based platform within the advertising ecosystem. To develop such tools where users effectively enforce various protection strategies, it require various important parameters of usability, flexibility, scalability etc., to be considered to give users transparency and control over their private data.

Another research direction would be to extend the analysis of privacy protection mechanisms to other different players, such as, advertisers, ad exchange, publishers with the aim to analyse and evaluate privacy policies and protection mechanisms that are claimed by these parties. This would help various entities in the advertising system to identify the flaws and further improve their working environment.

6.4 Private billing mechanism

Billing for both ad presentations and clicks is an important component of online advertising system. As discussed in

28. It [191] proposes a DNT field in the HTTP header that requests a web application to either disable the tracking (where it is automatically set) or cross-site the user tracking of an individual user.
Appendix B, a private billing proposal is based on Threshold BLS signature, Polynomial commitment, and Zero knowledge proof (ZKP), which are based on PIR mechanisms and Shamir secret sharing scheme along with Byzantine robustness. The applicability of this private billing model can be verified in the online advertising system, which would require changes on both the user and ad system side. Furthermore, note that the this private billing mechanism, implemented via polynomial commitment and zero-knowledge proof, is highly resource consuming process, henceforth, an alternative implementation with reduced processing time and query request size can be achieved via implementing together billing with PIR using multi-secret sharing scheme. In addition, to explore the effect of multi-secret sharing scheme in multiple-server PIR and hence comparative analysis to choose between the two variations of single-secret and multi-secret sharing system implementations. Multi-secret sharing scheme would help reduce the communication bandwidth and delays along with the processing time of query requests/responses.

In addition, our billing mechanism for ad presentations and clicks presented in [7], also described in Section 2.5, is applicable only to single ad requests with no impact on privacy. However, the broader parameter values (simultaneously processing multiple ad requests) and the use of other PIR techniques, such as Hybrid-PIR [115] and Heterogeneous-PIR [196], can be used to efficiently make use of processing time.

Furthermore, with the rise in popularity of Cryptocurrencies, many businesses and individuals have started investing in them, henceforth, the applicability of embedding the Cryptocurrency with the existing billing methods needs an investigation and developing new frameworks for coexisting the billing payments with the Cryptocurrency market. In addition, this would require techniques for purchasing, selling, and transferring Cryptocurrency among various parties i.e. ad systems, app developers, publishers, advertisers, crypto-markets, and miners. A further analysis would require investigating the impact of such proposals on the current advertising business model with/without a significant effect.

An important research direction is to explore implementation of private advertising systems in Blockchain networks since there is limited Blockchain-based advertising systems e.g., [28], [35]. The [28] presents the design of a decentralised framework for targeted ads that enables private delivery of ads to users whose behavioral profiles accurately match the presented ads, defined by the advertising systems. This framework provides: a private profiling mechanism, privately requesting ads from the advertising system, the billing mechanisms for ads monetisation, uploading ads to the cloud system, various types of transactions to enable advertising operations in Blockchain-based network, and access policy over cloud system for accessing various resources (such as ads, mobile user profiles). However, its applicability in an actual environment is still questionable, in addition to, the coexistence of ads-billing mechanism with Cryptocurrency.

7 Conclusion

Targeted/Online advertising has become ubiquitous on the internet, which has triggered the creation of new internet ecosystems whose intermediate components have access to billions of users and to their private data. The lack of transparency of online advertising, the A&A companies and their operations have posed serious risks to user privacy. In this article, we break down the various instances of targeted advertising, their advanced and intrusive tracking capabilities, the privacy risks from the information flow among various advertising platforms and ad/analytics companies, the profiling process based on user’s private data and the targeted ads delivery process. Several solutions have been offered in the literature to help protect user privacy in such a complex ecosystem, henceforth, we provide a wide range of mechanisms that were classified based on the privacy mechanisms used, ad serving paradigm and the deployment scenarios (browser and mobile). Some of the solutions are very popular among internet users, such as blocking, however their blocking mechanism negatively impacts the advertising systems. On the other hand, majority of the proposals provide naive privacy that require a lot of efforts from the users; similarly, other solutions demand structural changes with the advertising ecosystems. We have found that it is very hard, based on various privacy preserving approaches, while demanding for devising novel approaches, to provide user privacy that could give users more control over their private data and to reduce the financial impact of new systems without significantly changing the advertising ecosystems and their operations.

Appendix A

Private Information Retrieval (PIR)

PIR [110], [111], [115], [155], [156], [157] is a multiparty cryptographic protocol that allows users to retrieve an item from the database without revealing any information to the database server about the retrieved item(s). In one of our previous works [7], our motivation for using PIR rather than other solutions, e.g., Oblivious Transfer [178], [179], is the lower communication and computation overheads of such schemes.

A user wishes to privately retrieve \( j^{th} \) record(s) from the database \( D \). \( D \) is structured as \( r \times s \), where \( r \) is the number of records, \( s \) the size of each record; \( s \) may be divided into words of size \( w \). For multi-server PIR, a scheme uses \( l \) database servers and has a privacy level of \( t \); \( k \) is the number of servers that respond to the client’s query, among those, there are \( v \) Byzantine servers (i.e., malicious servers that respond incorrectly) and \( h \) honest servers that send a correct response to the client’s query. Following, we briefly discuss and compare various PIR schemes.
A.1 Computational PIR (CPIR)

The single-server PIR schemes, such as CPIR [109], rely on the computational complexity (under the assumption that an adversary has limited resources) to ensure privacy against malicious adversaries. To privately retrieve the $\beta^{th}$ record from $D$, a CPIR client creates a matrix $M_\beta$ by adding hard noise (based on large disturbance by replacing each diagonal term in $M_\beta$ by a random bit of $2^{|\text{th}}$ words) to the desired record and soft noise (based on small disturbance) to all the other records. The client assumes that the server cannot distinguish between the matrices with hard and soft noises. The server multiplies the query matrix $M_\beta$ to the database $D$ that results in corresponding response $R_i$; the client removes the noise from $R$ to derive the requested record $\beta^{th}$.

A.2 Recursive PIR (R-PIR)

The CPIR mechanism is further improved in terms of communication costs [109] by recursively using the single-server CPIR where the database is split into a set of virtual small record sets each considered as a virtual database. The query is hence calculated against part of the database during each recursion. The client recursively queries for the virtual records, each recursion results in a virtual database of smaller virtual records, until it determines a single (actual) record that is finally sent to the client.

A.3 Information Theoretic PIR (IT-PIR)

The multi-server IT-PIR schemes [27], [111], [112], [113], [114] rely on multiple servers to guarantee privacy against colluding adversaries (that have unbounded processing power) and additionally provide Byzantine robustness against malicious servers.

To query a database for $\beta^{th}$ record with protection against up to $t$ colluding servers, the client first creates a vector $e_\beta$, with ‘1’ in the $\beta^{th}$ position and ‘0’ elsewhere. The client then generates $(l, t)$ Shamir secret shares $v_1, v_2, \ldots, v_l$ for $e_\beta$. The shares (one each) are subsequently distributed to the servers. Each server $i$ computes the response as $R_i = v_i \cdot D$, this is sent back to the client. The client reconstructs the requested $\beta^{th}$ record of the database from these responses. The use of Shamir secret sharing enables the recovery of the desired record from (only) $k \leq l$ server responses [111], where $k > t$ (and $t < l$).

A.4 Hybrid-PIR (H-PIR)

The multi-server H-PIR scheme [115] combines multi-server IT-PIR [111] with the recursive nature of the single-server CPIR [109] to improve performance, by lowering the computation and communication costs. Let these two schemes be respectively represented by $\tau$ for IT-PIR and the $\gamma$ for the recursive CPIR protocol. A client wants to retrieve $\beta^{th}$ record then the client must determine the index of virtual records containing the desired records at each step of the recursion until the recursive depth $d$. The client creates an IT-PIR $\tau$-query for the first index and sends it to each server. It then creates CPIR $\gamma$-query during each of the recursive steps and sends it to all the servers. Similarly, on the server side at each recursive steps; the server splits the database into virtual records each containing actual records, uses the $\tau$ server computation algorithm, and finally uses the $\gamma$ CPIR server computation algorithm. The last recursive step results in the record $R_d$, that is sent back to the client.

A.5 Comparison and applicability of various PIR techniques in ad systems

Following comparative analysis, based on literature work, would help the selection of various PIR schemes and their applicability within an advertising system. We note that various performance metrics relate to the size of query along with the selection of a particular PIR scheme e.g., the CPIR takes longer processing delays and highest bandwidth consumption compared to both the IT-PIR and H-PIR schemes. This is due to the computations involved in query encoding and due to the servers performing matrix-by-matrix computations instead of vector-by-matrix, as is used by the IT-PIR and H-PIR schemes [115], although, the communication cost can be lowered down using the recursive version of the CPIR [109].

Furthermore, IT-PIR provides some other improvements, such as the robustness, which is its ability to retrieve correct records even if some of the servers do not respond or reply with incorrect or malicious responses [114]. It is further evident [115] that both the single-server CPIR and the multi-server IT-PIR schemes, such as [27], [111], [112], [113], respectively make the assumptions of computationally bounded and that particular thresholds of the servers are not colluding to discover the contents of a client’s query. Alternatively, the H-PIR [115], provides improved performance by combining multi-server IT-PIR with the recursive nature of single-server CPIR schemes respectively to improve the computation and communication costs.

A recent implementation i.e., Heterogeneous PIR [196], enables multi-server PIR protocols (implemented using multi-secret sharing algorithm, compatible with Percy++ [30] PIR library) over non-uniform servers (in a heterogeneous environment where servers are equipped with diverse resources e.g., computational capabilities) that impose different computation and communication overheads. This implementation makes it possible to run PIR over a range of different applications e.g., various resources (ad’s contents such as, JPEG, JavaScript

29. A complete implementation of CPIR, IT-PIR and H-PIR, Percy++ is present on http://percy.sourceforge.net/.

30. http://percy.sourceforge.net/
files) present on CDN in distributed environments. Furthermore, this implementation has tested and compared its performance with Goldberg’s [111] implementation with different settings e.g., for different database sizes, numbers of queries and for various degrees of heterogeneity. This implementation achieves a trade-off between computation and communication overheads in heterogeneous server implementation by adjusting various parameters.

APPENDIX B
BUILDING BLOCKS FOR ENABLING PIR AND PRIVATE BILLING

This section introduces various building blocks for enabling PIR techniques i.e. Shamir secret sharing and Byzantine robustness. It further discusses various techniques that are used for private billing i.e. Threshold BLS signature, Polynomial commitment, and Zero-knowledge proof (ZKP).

B.1 Shamir secret sharing

The Shamir secret sharing [197] scheme divides a secret \( \sigma \) into parts, giving each participant e.g. \( l \) servers a unique part where some or all of the parts are needed in order to reconstruct the secret. If the secret is found incorrect then it can be handled through error-correcting codes, such as the one discussed in [198]. Let the \( \sigma \) be an element of some finite field \( F \) then the Shamir scheme works as follows: a client selects an \( l \) distinct non-zero elements \( \alpha_1, \alpha_2, \cdots, \alpha_l \in F \) and selects \( t \) elements \( a_1, a_2, \cdots, a_t \in \mathbb{F} \) (the \( \mathbb{F} \) means uniformly at random). A polynomial \( f(x) = \sigma + a_1x + a_2x^2 + \cdots + a_t x^t \) is constructed and gives the share \( (\alpha_i, f(\alpha_i)) \in F \times F \) to the server \( i \) for \( 1 \leq i \leq l \). Now any \( t+1 \) or more servers can use Lagrange interpolation [114] to reconstruct the polynomial \( f \) and, similarly, obtains \( \sigma \) by evaluating \( f(0) \).

B.2 Byzantine robustness

The problem of Byzantine failure allows a server to continue its operation but it incorrectly responds. The Byzantine failure may include corrupting of messages, forging messages, or sending conflicting messages through malicious or errors. In order to ensure the responses’ integrity in a single-server, such as PIR-Tor [199], the server can provide a cryptographic signature on each database’s block. However, in a multi-server PIR environment, the main aim of the Byzantine robustness is to ensure that the protocol still functions correctly even if some of the servers fail to respond or provide incorrect or malicious responses. The client at the same time might also be interested in figuring out which servers have sent incorrect responses so that they can be avoided in the future.

The Byzantine robustness for PIR was first considered by Beimel and Stahl [200], [201]; the scheme called the \( t \)-private \( v \)-Byzantine robust \( k \)-out-of-\( l \) PIR. The authors take the \( l \)-server information-theoretic PIR setting where \( k \) of the servers respond, \( v \) servers respond incorrectly, and the system can sustain up to \( t \) colluding servers without revealing client’s query among them. Furthermore, they suggest the unique decoding where the protocol always outputs a correct unique block under the conditions \( v \leq t \leq k/3 \).

The [111] uses the list decoding, that is an alternative to unique decoding of error-correcting codes for large error rates, and demonstrates that the privacy level can be substantially increased up to \( 0 < t < k \) and the protocol can tolerate up to \( k - \lfloor \sqrt{k/3} \rfloor - 1 \) Byzantine servers. Alternatively, the list decoding can also be converted to unique decoding [202] at the cost of slightly increasing the database size [114].

Following schemes are the essential building blocks for enabling private billing along with evaluating the PIR techniques for privately retrieving ads from the ad database.

B.3 Threshold BLS signature

The Boneh-Lynn-Shacham (BLS) [203] is a ‘short’ signature verification scheme that allows a user to verify that the signer is authentic. The signer’s private signing key is a random integer \( x \in \mathbb{Z}_q \) and the corresponding public verification key is \( (g^x, g) \) (\( g \) is a generator of \( \mathbb{G}_2 \)). The procedure for signature verification is as follows: Given the signing key \( x \) and a message \( m \), the signature is computed via \( \sigma = h_x \) where \( h = \text{hash}(m) \) is a cryptographic hash of \( m \); the verification equation is \( \epsilon(\sigma, \hat{g}) = \epsilon(h, \hat{g}^x) \), which results in true/false. To fit into scenario of multiple PIR servers; a \( (k,l) \)-threshold variant of BLS signature can be used where signing keys are the evaluations of a polynomial of degree \( (k-l) \) and the master secret is the constant term of this polynomial. Similarly, the reconstruction process can be done using Lagrange interpolation. The \( (k-l) \) threshold BLS signature partly provides the level of robustness against the Byzantine signers since the signature share can be verified independently using the signer’s public verification key share.

B.4 Polynomial commitment

A polynomial commitment [204] scheme allows committers to formulate a constant-sized commitments to polynomials that (s)he can commit so that it can be used by a verifier to confirm the stated evaluations of the committed polynomial [205], without revealing any additional information about the committed value(s). An example of the Polynomial commitment constructions in [204] provides unconditional hiding if a commitment is opened to at most \( t-1 \) evaluations (i.e. \( t-1 \) servers for a degree-\( t \) polynomial) and provides computational hiding under the discrete \( \log(D_t) \) if polynomial commitment is opened to at least \( t \) evaluations. As presented in [204],
commitment to a polynomial \( f(x) = a_1x^1 + \cdots + a_1z + a_0 \) has the form 
\[
C_{f} = (\alpha^{g^{a_1}} \cdots \alpha^{g^{a_1}} g^{a_0}) = f^{(\alpha)}
\]
where \( \alpha \) is secret, \( g \in G_1 \) is a generator whose discrete logarithm with respect to \( g \) is unknown, including all the bases are part of the commitment scheme’s public key. The verifier, on the other side, can confirm that the claimed value is \( g \) is.

B.5 Zero-knowledge proof (ZKP)

The zero knowledge proof is an interactive protocol between the prover and the verifier that allows the prover to prove to the verifier that it holds a given statement without revealing any other information. There are several ZKPs, such as range proof to prove that a committed value \( \alpha \) is true, here the commitment \( w \) is called the witness; detailed discussion can be found in [204].

APPENDIX C

K-ANONYMITY

k-anonymity was introduced in [105], [208] and its enforcement through generalization and suppression was suggested in [106]. k-anonymity examines the identification attack, which aims to release private vector of the data (i.e. structured data e.g. data holders of bank or hospital etc.) that cannot be re-identified while the data still remains useful. Let \( RT (A_1, \ldots, A_n) \) be a set of structured data organised in rows and columns, a population of entities \( U \), with a finite set of attributes of \( RT \) as \( (A_1, \ldots, A_n) \) with at least one attribute identified as ‘key attribute’ that can be considered as quasi-identifier[31, 32]. A quasi-identifier of \( RT \), represented as \( Q_{RT} \), is a set of attributes \( (A_1, \ldots, A_j) \subseteq (A_1, \ldots, A_n) \), where \( \exists p_i \subseteq U \) such that \( f_p: f_q: \frac{Q_{RT}}{p_i}; f_c: U \to RT \) and \( f_q: RT \to U^*, U \subseteq U^* \).

k-anonymity for \( RT \) is achieved if each sequence of values in \( RT [Q_{RT}] \) appears with at least \( k \) occurrences i.e. \( Q_{RT} = (A_1, \ldots, A_j) \) be the quasi-identifier associated with \( RT \), where \( A_1, \ldots, A_j \subseteq A_1, \ldots, A_n \), and \( RT \) satisfy k-anonymity. Subsequently, each sequence of values in \( RT [Q_{RT}] \) appears with at least \( k \) occurrences in \( RT [Q_{RT}] \).

31. Variable values or combinations of variable values within a dataset that are not structural unique but might be empirically unique and therefore in principle uniquely identify a population unit. https://stats.oecd.org/glossary/detail.asp?ID=8961
32. Quasi-identifiers are pieces of information that are not of themselves unique identifiers, but are sufficiently well correlated with an entity that they can be combined with other quasi-identifiers to create a unique identifier. https://en.wikipedia.org/wiki/Quasi-identifier for \( x = i, \ldots, j \). The \( RT \) satisfies the k-anonymity is released. The combination of any set of attributes of the released data \( RT \) and external sources on which \( Q_{RT} \) (PT is the private table) is based, cannot be linked that eventually guarantees the privacy of released data. A detailed example is given in [105].

REFERENCES

[1] GreensMedia, “45 digital and targeted advertising statistics,” https://www.greensmedia.com/blog/45-digital-and-targeted-advertising-statistics/.
[2] buildfire, “Mobile app download and usage statistics (2020),” https://buildfire.com/app-statistics/.
[3] M. C. Grace, W. Zhou, X. Jiang, and A.-R. Sadeghi, “Unsafe exposure analysis of mobile in-app advertisements,” pp. 101–112, 2012.
[4] T. Book and D. S. Wallach, “A case of collusion: A study of the interface between ad libraries and their apps,” pp. 79–86, 2013.
[5] A. Chaabane, G. Acs, and M. A. Kaafar, “You are what you like! Information leakage through users’ interests,” 2012.
[6] C. Castelluccia, M.-A. Kaafar, and M.-D. Tran, “Betrayed by your ads!,” pp. 1–17, 2012.
[7] I. Ullah, B. G. Sarwar, R. Boreli, S. S. Kanhere, S. Katenbeisser, and M. Hollick, “Enabling privacy preserving mobile advertising via private information retrieval,” in 2017 IEEE 42nd Conference on Local Computer Networks (LCN), pp. 347–355, IEEE, 2017.
[8] I. Ullah, R. Boreli, S. S. Kanhere, S. Chawla, T. A. Ahanger, and U. Thaqir, “Protecting privacy and attributes in app based mobile user profiling,” IEEE Access, vol. 8, pp. 143818–143836, 2020.
[9] T. Chen, I. Ullah, M. A. Kaafar, and R. Boreli, “Information leakage through mobile analytics services,” in 15th International Workshop on Mobile Computing Systems and Applications, ACM HotMobile, 2014.
[10] S. Mamais, Privacy-preserving and fraud-resistant targeted advertising for mobile devices. PhD thesis, Cardiff University, 2019.
[11] Y. Liu and A. Simpson, “Privacy-preserving targeted mobile advertising: requirements, design and a prototype implementation,” Software: Practice and Experience, vol. 46, no. 12, pp. 1657–1684, 2016.
[12] Y. Wang, E. Genc, and G. Peng, “Aiming the mobile targets in a cross-cultural context: Effects of trust, privacy concerns, and attitude,” International Journal of Human–Computer Interaction, vol. 36, no. 3, pp. 227–238, 2020.
[13] CNET, “Facebook vs. apple: Here’s what you need to know about their privacy feud,” https://www.cnet.com/news/facebook-vs-apple-heres-what-you-need-to-know-about-their-privacy-feud/.
[14] I. Ullah, M. A. Kaafar, and S. Katenbeisser, “Don’t kill my ads!: balancing privacy in an ad-supported mobile application market,” p. 2, 2012.
[15] N. Vallina-Rodriguez, J. Shah, A. Finanmore, Y. Grunenberger, K. Papagiannaki, H. Haddadi, and J. Crowcroft, “Breaking for commercials: characterizing mobile advertising,” pp. 343–356, 2012.
[16] S. Han, J. Jung, and D. Wetherall, “A study of third-party tracking by mobile apps in the wild,” 2012.
[17] “Flurry advertisers, publishers, and analytics,” www.flurry.com.
[18] I. Ullah, R. Boreli, M. A. Kaafar, and S. S. Kanhere, “Characterizing user targeting for in-app mobile ads,” pp. 547–552, 2014.
[19] “Mobile advertising market size, share & industry analysis, forecast 2019-2026,” https://www.fortunebusinessinsights.com/mobile-advertising-market-102486, Accessed on June, 2020.
[20] V. Ng and M. K. Ho, “An intelligent agent for web advertisements,” International Journal of Foundations of Computer Science, vol. 13, no. 04, pp. 531–544, 2002.
A. Thawani, S. Copalan, and V. Sridhar, “Event driven semantics based ad selection,” vol. 3, pp. 1875–1878, 2004.
[21] J. Yan, N. Liu, G. Wang, W. Zhang, Y. Jiang, and Z. Chen, “How much can behavioral targeting help online advertising?,” pp. 261–270, 2009.
[22] J. Jaworska and M. Sydow, “Behavioural targeting in on-line advertising: An empirical study,” in Web Information Systems Engineering—WISE 2008., pp. 62–76, Springer, 2008.
[23] J. Shin and J. Yu, “Targeted advertising: How do consumers make inferences?,” 2019.
A. Mukherjee and B. Liu, "Improving gender classification of blog authors," in *Proceedings of the 2010 conference on Empirical Methods in natural language Processing*, pp. 207–217, Association for Computational Linguistics, 2010.

B. Bi, M. Shokouhi, M. Kosinski, and T. Graepel, "Inferring the demographics of search users: social data meets search queries," in *22nd International conference on Web search and Data mining*, pp. 131–140, 2013.

J. J.-C. Ying, Y.-J. Chang, C.-M. Huang, and V. S. Tseng, "Demographic prediction based on users mobile behaviors," *Mobile Data Challenge*, 2012.

J. W. Pennebaker, M. E. Francis, and R. J. Booth, "Linguistic inquiry and word count: LIWC 2001," *Mahw: Lawrence Erlbaum Associates*, vol. 71, p. 2001.

E. Zheleva and L. Getoor, "To join or not to join: the illusion of privacy in social networks with mixed public and private user profiles," pp. 351–540, 2009.

J. He, W. W. Chu, and Z. V. Liu, "Inferring privacy information from social networks," pp. 154–165, 2006.

A. Mislove, B. Viswanath, K. P. Gummadi, and P. Druschel, "You are who you know: inferring user profiles in online social networks," pp. 251–260, 2010.

E. Ryu, Y. Rong, J. Li, and A. Machanavajjhala, "course: protect data from course of attribute inference: a social network privacy-analyzer," in *Proceedings of the ACM SIGMOD Workshop on Databases and Social Networks*, pp. 13–18, ACM, 2013.

W. Enck, P. Gilbert, S. Han, V. Tendulkar, B.-G. Chun, L. P. Cox, J. Jung, P. McDaniel, and A. N. Sheth, "Taintdroid: an information-flow tracking system for realtime privacy monitoring on smartphones," *ACM Transactions on Computer Systems* (TOCS), vol. 32, no. 2, p. 5, 2014.

M. Ongtang, S. McLaughlin, W. Enck, and P. McDaniel, "Seminically rich application-centric security in android," *Security and Communication Networks*, vol. 5, no. 6, pp. 658–673, 2012.

A. Frik, A. Haviland, and A. Acquisti, "The impact of ad-blockers on product search and purchase behavior: A lab experiment," in *Proceedings of the USENIX Security Symposium* (USENIX Security 20), 2020.

A. Shuba and A. Markopoulou, "Nomoats: Towards automatic detection of mobile tracking," in *Proceedings of Privacy Enhancing Technologies*, vol. 2, pp. 45–66, 2020.

U. Iqbal, P. Snyder, S. Zhu, B. Livshits, Z. Qian, and Z. Shafiq, "Adgraph: A graph-based approach to ad and tracker blocking," in *Proc. of IEEE Symposium on Security and Privacy*., 2020.

A. P. Felt, E. S. Egelman, A. C. D. Wagner, "Android permissions: User attention, comprehension, and behavior," in *Proceedings of the eighth symposium on usability and privacy security*, pp. 1–12, 2014.

A. P. Felt, H. J. Wang, A. Moshchuk, S. Hanna, and E. Chin, "Permission Re-Delegation: Attacks and Defenses," 2011.

A. P. Felt, E. Chin, S. Hanna, D. Song, and D. Wagner, "Android permission satisfiability," in *Proceedings of the 18th ACM conference on Computer and communications security*, pp. 627–638, 2011.

P. P. Chan, L. C. Hui, and S.-M. Yiu, "Droidchecker: analyzing android applications for capability leak," in *Proceedings of the fifth ACM conference on Security and Privacy in Wireless and Mobile Networks*, pp. 125–136, 2012.

W. Enck, M. Ongtang, and P. McDaniel, "On lightweight mobile phone application certification," in *Proceedings of the 16th ACM conference on Computer and communications security*, pp. 235–245, 2009.

A. R. Beresford, A. Rice, N. Shehin, and R. Sohan, "Mockdroid: trading privacy for application functionality on smartphones," pp. 49–54, 2011.

P. Hornyack, S. Han, J. Jung, S. Schechter, and D. Wetherall, "These aren’t the droids you’re looking for: retrofitting android to protect data from impious applications," in *Proceedings of the 18th ACM conference on Computer and communications security*, pp. 639–652, 2011.

P. Golle and K. Partridge, "On the anonymity of home/work location pairs," in *International Conference on Pervasive Computing*, pp. 390–397, Springer, 2009.

H. Zang and J. Bolot, "Anonymization of location data does not work: A large-scale measurement study," in *Proceedings of the 17th annual international conference on Mobile computing and networking*, pp. 145–156, 2011.

N. Mohammed, B. C. Fung, and M. Debbabi, "Walking in the crowd: anonymizing trajectory data for pattern analysis," in *Proceedings of the 18th ACM conference on Information and knowledge management*, pp. 1441–1444, 2009.

F. Bonchi, L. V. Lakshmanan, and H. Wang, "Trajectory anonymity in publishing personal mobility data," *ACM Sigkdd Explorations Newsletter*, vol. 13, no. 1, pp. 30–40, 2011.

R. Shokri, G. Theodorakopoulos, G. Danezis, J.-P. Hubaux, and J.-Y. Le Boudec, "Quantifying location privacy: the case of sporadic location exposure," in *International Symposium on Privacy Enhancing Technologies Symposium*, pp. 57–76, Springer, 2011.

P. Samarati, "Protecting respondents identities in microdata release," *IEEE transactions on Knowledge and Data Engineering*, vol. 13, no. 6, pp. 1010–1027, 2001.

L. Sweeney, "k-anonymity: A model for protecting privacy," *International Journal of Uncertainty, Fuzziness and Knowledge-Based Systems*, vol. 10, no. 05, pp. 557–570, 2002.

A. Machanavajjhala, D. Kifer, J. Gehrke, and M. Venkatasubramaniam, "l-diversity: Privacy beyond k-anonymity," *ACM Transactions on Knowledge Discovery from Data* (TKDD), vol. 1, no. 1, p. 3, 2007.

N. Li, T. Li, and S. Venkatasubramanian, “t-closeness: Privacy beyond k-anonymity and l-diversity,” pp. 106–115, 2007.

C. Aguilar Melchor and P. Gaborit, “A lattice based computationally efficient private information retrieval protocol,” pp. 446, 2007.

B. Chor and N. Gilboa, "A computationaly private information retrieval protocol," pp. 304–313, 1997.

I. Goldberg, "Improving the robustness of private information retrieval," pp. 131–148, 2007.

A. Beimel, Y. Ishai, and T. Malkin, “Reducing the servers computation in private information retrieval: Pir with preprocessing,” *Journal of Cryptology*, vol. 17, no. 2, pp. 125–151, 2004.

Y. Gertner, S. Goldwasser, and T. Malkin, "A random server model for private information retrieval," pp. 200–217, 1998.

R. Shokri, N. H. Devet, I. Goldberg, "Computationally robust private information retrieval,” pp. 269–283, 2012.

C. Devet and I. Goldberg, "The best of both worlds: Combining information-theoretic and computational pir for communication efficiency," pp. 63–82, 2014.

M. Fredrikson and B. Livshits, "Repr: Re-imaging content personalization and in-browser privacy," pp. 131–146, 2011.

S. Guha, B. Cheng, and P. Francis, "Privad: Practical privacy in online advertising.," pp. 20–30, 2011.

R. Chen, A. Reznichenko, P. Francis, and J. Gehrke, "Towards statistical queries over distributed private user data," in *Presented as part of the 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 12)*, pp. 169–182, 2012.

R. Chen, I. E. Akkus, and P. Francis, "Splits: high-performance private analytics," pp. 315–326, 2013.

M. M. Tsang, S.-C. Ho, and T.-P. Liang, "Consumer attitudes toward mobile advertising: An empirical study," *International journal of electronic commerce*, vol. 8, no. 3, pp. 65–78, 2004.

M. Merisavo, S. Kajalo, H. Karjaluoto, V. Virtanen, S. Salmenkivi, M. Raulas, and M. Leppäniemi, "An empirical study of the drivers of consumer acceptance of mobile advertising," *Journal of interactive advertising*, vol. 7, no. 2, pp. 41–50, 2007.

G. A. Johnson, S. K. Shriver, and S. Du, "Consumer privacy choice in online advertising: Who opts out and at what cost to industry?," *Marketing Science*, 2020.

R. Dingleline, N. Mathewson, and P. Syverson, "Tor: The second-generation onion router," 2004.

G. Aggarwal, E. Bursztein, C. Jackson, and D. Boneh, "An analysis of private browsing modes in modern browsers," pp. 79–94, 2010.

I. E. Akkus, R. Chen, M. Hardt, P. Francis, and J. Gehrke, "Non-tracking web analytics," 2012.

M. Backes, A. Kate, M. Maffei, and K. Pecina, "Obliviad: Provably secure and practical online behavioral advertising," pp. 257–271, 2012.

M. Hardt and S. Nath, "Privacy-aware personalization for mobile advertising," 2012.

P. Samarati and L. Sweeney, "Generalizing data to provide anonymity when disclosing information," in *PODS*, vol. 98, p. 188, 1998.
C. Riederer, V. Erramilli, A. Chaintreau, B. Krishnamurthy, and A. Ghosh and A. Roth, “Selling privacy at auction,” Games and S. Christopher, S. Sid, and K. Dan I. E. Akkus, R. Chen, M. Hardt, P. Francis, and J. Gehrke, “Non- J. Freudiger, N. Vratonjic, and J.-P. Hubaux, “Towards privacy- A. Dorri, M. Steger, S. S. Kanhere, and R. Jurdak, “Blockchain: A distributed solution to automotive security and privacy,” IEEE Communications Magazine, vol. 55, no. 12, pp. 119–125, 2017. S. Nakamoto, “Bitcoin: A peer-to-peer electronic cash system,” tech. rep., Manubot, 2019. V. Dedeoglu, R. Jurdak, A. Dorri, R. Lunardi, R. Michelin, A. Zorzo, and S. Kanhere, “Blockchain technologies for IoT,” in Advanced Applications of Blockchain Technology, pp. 53–89, Springer, 2020. J. Yang, J. Wen, B. Jiang, and H. Wang, “Blockchain-based sharing and tamper-proof framework of big data networking,” IEEE Network, vol. 34, no. 4, pp. 62–67, 2020. A. Tandon, A. Dhir, N. Islam, and M. M¨antym¨aki, “Blockchain for cyberphysical systems including private advertising and crypto-based billing systems. His primary research interest include privacy enhancing technologies; he also has interest in Internet of Things, Blockchain, network modeling and design, network measurements, and trusted networking.

Roksana Boreli has received her Ph.D in Communications from University of Technology, Sydney, Australia. She has over 20 years of experience in communications and networking research and in engineering development, in large telecommunications companies (Telstra Australia, Xantic, NL) and research organisations. Roksana has served in various positions of Engineering manager, Technology strategist, Research leader of the Privacy area of Networks research group in National ICT Australia (NICTA)/CSIRO Data61 and CTO in a NICTA spinoff 7-ip. Her primary research focus is on the privacy enhancing technologies; she also maintains an interest in mobile and wireless communications.

Salil S. Kanhere (Senior Member, IEEE) received the M.S. and Ph.D. degrees from Drexel University, Philadelphia. He is currently a Professor of Computer Science and Engineering with UNSW Sydney, Australia. His research interests include the Internet of Things, cyberphysical systems, blockchain, pervasive computing, cybersecurity, and applied machine learning. He is a Senior Member of the ACM, an Humboldt Research Fellow, and an ACM Distinguished Speaker. He serves as the Editor in Chief of the Ad Hoc Networks Journal and as an Associate Editor of the IEEE Transactions On Network and Service Management, Computer Communications, and Pervasive andMobile Computing. He has served on the organising committee of several IEEE/ACM international conferences. He has co-authored a book titled Blockchain for Cyberphysical Systems.