Chapter 9
Tracking and Personalization

Rahat Masood, Shlomo Berkovsky, and Mohamed Ali Kaafar

Abstract This chapter studies the relationship between two important, often conflicting paradigms of online services: personalization and tracking. The chapter initially focuses on the categories and levels of online personalization, briefly overviewing algorithmic methods applied to achieve these. Then, the chapter turns to online tracking specific to mobile and web technologies, as well as the more advanced behavioral tracking. Following this, the chapter ties the streams of personalization and tracking together and discusses various aspects of their relationships, including the currently deployed tracking methods for personalization. Privacy implications of personalization via online tracking, highlighted by organizations and researchers, are also illustrated. Lastly, this chapter discusses the ways to balance personalization benefits and privacy concerns. This includes the state-of-the-art practices, current challenges, and practical recommendations for system developers willing to strike this balance.

9.1 Introduction

The ever-changing technological landscape, high user involvement, increased societal visibility, and amalgamation of services have made privacy challenging to maintain in a digital world. In recent years, we have witnessed many privacy violation incidents where tech-giant companies (e.g., Google, Facebook, LinkedIn) were involved. For instance, at Princeton University, computer-science researchers

R. Masood (✉)
Data61-CSIRO, University of New South Wales (UNSW), Sydney, NSW, Australia
e-mail: rahat.masood@data61.csiro.au

S. Berkovsky
Macquarie University, Sydney, NSW, Australia
e-mail: shlomo.berkovsky@mq.edu.au

M. A. Kaafar
Macquarie University and Data61-CSIRO, Sydney, NSW, Australia
e-mail: dali.kaafar@mq.edu.au

B. P. Knijnenburg et al. (eds.), *Modern Socio-Technical Perspectives on Privacy*,
https://doi.org/10.1007/978-3-030-82786-1_9

confirmed that Google services on Android devices and iPhones store users' location data, even if users set their privacy settings to prevent Google from (geo)locating a user [1]. Similarly, Facebook has often been involved in scandals such as the Cambridge Analytica data harvesting [2], suspicions of Russian and Iranian meddling in the US elections [3], and several data-exposing "bugs" [4]. According to [5], roughly 17,000 Android apps collect identifying information about a user by setting persistent identifiers on mobile phones. These identifiers are the unique numbers that allow companies to learn about user's activities on a mobile phone. These examples indicate that a growing number of service providers use several techniques to collect a wide variety of data about end-users, from basic socio-demographic details to a complete history of a user's searches, clicks, locations, and details of the device used.

One apparent reason to collect such information is to create a personalized user experience to increase revenue, but at the same time, this information may also be used for different purposes, such as to build user profiles to strengthen user engagement and loyalty. Moreover, in some cases, this information may be shared with third parties to assist in various tasks such as sharing on social platforms, hosting and maintenance, or customer care [6]. The plethora of cases where companies collect as much information about end users as possible, sometimes unknowingly to them, and then using it for personalization, has raised the awareness of various issues associated with the need to preserve and maintain users' privacy. This chapter details the various aspects of the relationship between online tracking and personalization, including the currently deployed tracking methods for personalization, and existing solutions to balance personalization benefits and privacy concerns.

9.2 Aspects of Personalization

Personalized technologies are deployed nowadays by virtually every website and mobile app. These technologies facilitate the "provision of content and services tailored to individuals based on knowledge about their preferences and behavior" [7]. While personalized services started two decades ago with use cases like web content filtering and eCommerce recommendations, they have since spread to applications like music, tourism, eHealth, and more [8]. In this subsection we initially overview the goals and benefits of personalized technologies and then discuss their applications in the web and mobile environments.

9.2.1 Goals of Personalization

Naturally, the tailoring of services offered by personalization can benefit both the service provider and the end user. For the former, it allows to increase the quality of the service, as it gets adjusted to the needs and preferences of the user. This can

lead to tangible improvements in various metrics, such as user engagement, click-through rate, returning users, positive feedback, and, in consequence, to increased revenues for the service. Likewise, users also benefit from the personalization, as the overall user experience is improved. For example, personalization can shorten the discovery of a desired content or reduce the costs of buying a product.

Many algorithmic approaches for personalization have been developed, evaluated, and deployed. Some of them rely on statistical correlations of past user behavior [9], while others capitalize on extensive domain knowledge [10]. Regardless of the underlying personalization algorithm, a necessary precondition for personalized services is the availability of reliable and up-to-date representation of the user, that is, their interests, preferences, and needs, as encapsulated by the user model [11].

User models typically reflect the goals and domain of the personalized service. For example, an email filtering plugin should be able to distinguish between genuine senders and spammers, while a movie recommender should know what movie genres are liked and disliked by the user. Thus, no one-size-fits-all representation of the user model can be conceived, and the target data is learned implicitly from observable user interactions with the system and other users.

Moreover, the information collected for personalizing the service is closely related to the underlying personalization algorithm. For example, collaborative methods rely on identifying similar users and deriving predictions for the target user from the behavior of the identified similar users. As such, collaborative methods naturally require knowledge about numerous users and the privacy concerns are harder to enforce in this case [12]. On the contrary, content-based methods require only the model of the target user and additional domain knowledge. The privacy of the latter is easier to protect than in the collaborative case, as the domain knowledge typically does not include any personal data [13].

9.2.2 Personalization Environments

The increasing use of mobile technologies has led to multi-modality (cross platforms), which allows users to access content and services through the web as well as through apps and mobile devices. In this section, we briefly describe these modalities with respect to personalization technologies.

9.2.2.1 Web Personalization

User modeling for web personalization purposes typically involves making sense of users' past information access and their interactions with online systems and other users. The facets of user data that can be modeled are diverse; for instance, they may include users' knowledge level, interests, goals and motivation, personality, and language. Potential sources of such user modeling data include past visited pages,

launched search queries, purchased products, played songs and video clips, friended social network users, liked content, and more [11].

Note that, if accessible, these sources allow the personalized services not only to populate the desired user model facets but also to derive additional sensitive information that may undermine users' privacy [14]. Depending on the richness and reliability of the user models, the service can either be truly personalized or just tailored, for example, according to the group to which the user belongs.

9.2.2.2 Mobile Personalization

The use case of mobile personalization adds another layer of information, often referred to as the *context*. The most prominent example of the contextual user model is users' location. This can be leveraged for a range of location-aware personalized services, such as recommendations of places of interest, weather and traffic alerts, presence of other people nearby, and so on. Within such services, either the user modeling data or the personalized options are preselected according to the user's current location [15].

Other available sources of mobile user modeling data are various sensors deployed by the mobile device. These include accelerometers and gyroscopes that track movement, biometric sensors that recognize faces as well as scan eye iris and fingerprints, light sensors that detect the ambient illumination level, microphones that can detect background noises, and more. Add to these the plethora of behavioral and interaction data that can potentially be extracted from the installed mobile apps, such as browsing logs, social media friends, physical activity data, commute and driving patterns, and so forth. In combination, mobile phones can collect a large variety of user data and allow constructing detailed user models [16].

Having obtained and processed this information, various adaptive services and suggestions can be tailored to the users' preferences and interests. For example, recommended retailers can be restricted to the user's current location [17], screen brightness can be adjusted to the ambient light intensity [18], timing and frequency of reminders can be tuned according to interactions with similar reminders [19], and driving route can be modified if traffic to the desired location is slow [20]. In the next section, we focus on tracking techniques and discuss in detail the relevant entities and mechanisms that facilitate personalization.

9.3 Online Tracking

Research has shown that desktops and mobile devices and associated web browsers and mobile apps contain subtle information that allows them to be "fingerprinted or tracked." Online tracking has several meanings, but one of the most valid general definition is "following the trails and movements of someone on the Internet through means such as mobile phones, desktop, and smart devices, in order to gain

unique information about them for incentives such as target advertising, profiling, and data exchange" [21]. Online tracking has various types and extensions: from detecting user interests when visiting a web page to recording various detailed aspects about the user, including their location, social relations, health, and political beliefs. A combination of such information increases the chances of identifying and appropriately tracking a user online. In Fig. 9.1, we show the ecosystem of online tracking. Additionally, the increasing use of IoT devices, such as SmartWatches, Fitbits, and SmartShoes, has made online tracking more aggravated as these devices collect, process, store, and disseminate sensitive users' data, such as health conditions, billing information, physical environment, and behavioral information. In Chap. 11, privacy issues in IoT devices are discussed in detail.

9.3.1 Tracking Contexts

There are several ways to achieve online tracking. In general, we contextualize them in four categories:

Web Tracking is one of the primary sources of the profiling that tracks users across different visits or sites. There are various design, implementation, and deployment methods that enable web tracking. For instance, for an externally hosted website, a service provider can embed third-party content or incorporate dynamic content like JavaScript snippets or libraries supplied by third party to implement the tracking functionality. In fact, more than 90% of Alexa's top-500 websites contain third-party tracking content [22], and that 70% of the cookies recorded were third-party cookies set by just 25 third-party domains [23]. That means the entities with whom the user may or may not have chosen to interact on the web may be recording their online behavior in unexpected ways.

Mobile Tracking identifies users through the devices equipped with sophisticated sensors, such as microphones, GPS, accelerometers, and cameras. These sensors generate highly sensitive data that can be used as unique fingerprints.[1] Like web tracking, mobile devices contain various identifiers that can be used (in isolation or in combination) to track or profile users. For example, researchers demonstrated how the use of WiFi SSID (the Service Set IDentifier representing the WiFi network devices connect to) in its active discovery mode could lead to revealing the geographical location of users [24] or distinguishing WiFi-enabled devices [25]. Others have shown how to infer the social relationship between mobile device owners by tracking their WiFi fingerprints [26]. Others have used motion sensor signals to identify devices or users [27–29]. The privacy concerns of mobile tracking are different from web tracking because of the diverse range of data available through sensors, apps, and mobile browsers. The high interconnectivity

[1] In privacy terminology, a fingerprint refers to a trace of information, often an observable characteristics of a device or a user, that is unique enough for identification or tracking purposes.

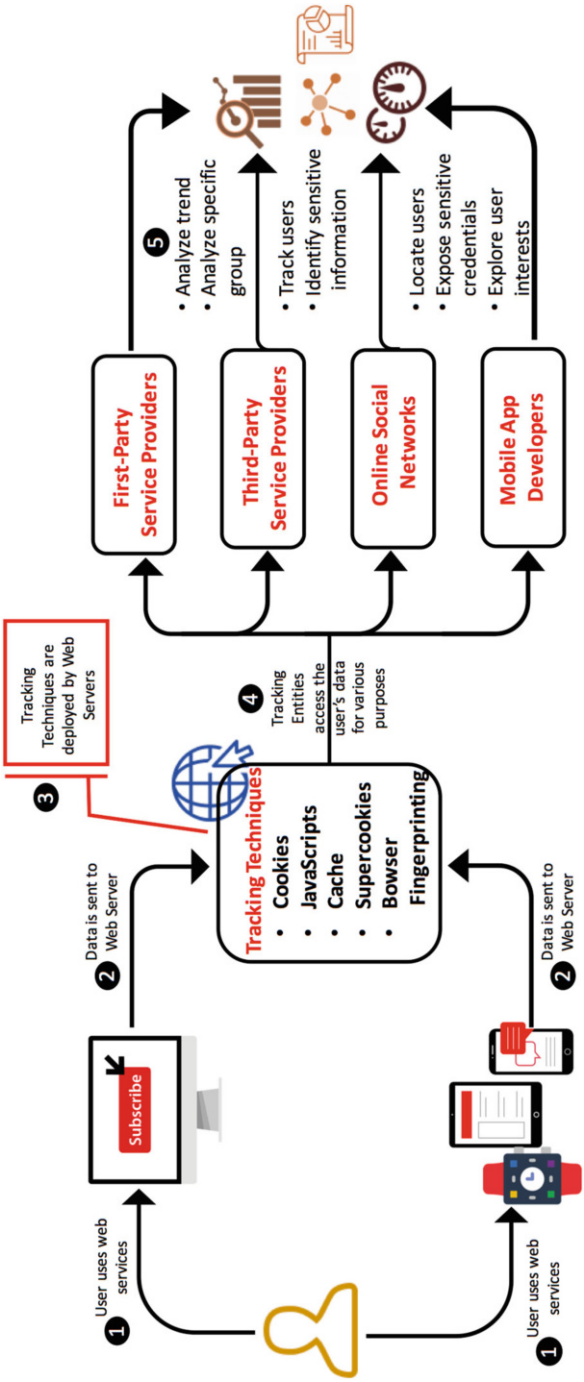

Fig. 9.1 Ecosystem of online tracking

and portability features of mobile devices have made them a perfect target for tracking.

Cross-Device Tracking is performed by many organizations today and can provide a more comprehensive view into users' behavior. There are several reasons to perform cross-device tracking. It allows consumers to log in to their email or social media accounts from multiple devices to maintain a "state" so they can pick up where they left off on a different device. It also facilitates companies to prevent fraud; for instance, if there is an unrecognized device, a company can take steps—such as sending an authentication code to an email address or phone number—to ensure that the new device belongs to the consumer who is trying to access an existing account. Companies also use cross-device tracking to improve user experience by personalizing the content on a website or an app and to accurately retarget a user on multiple devices by displaying relevant ads. Consider an example where a user searches for a movie ticket on a web browser of his desktop. He later used his mobile phone browser, which showed an advertisement of the same movie running in nearby cinemas.

Cross-App Tracking can be considered a particular form of cross-device tracking, where an app identifies other apps installed on the device and makes a link to a user [30]. For example, it was shown that user traits and whether or not the user is a parent of small children could be predicted from the installed mobile apps [31]. Similarly, a search of a discounted movie ticket on a Groupon app may result in ads for movies in theater on a Facebook app.

9.3.2 Tracking Entities

The abovementioned web and mobile tracking are used by *first-party* and *third-party* tracking entities, respectively. These entities perform tracking for purposes ranging from improved user experience to credit scoring or targeted political messages. We explain these two types of tracking below.

9.3.2.1 First-Party Tracking

First-party tracking is performed by the service providers with which the user interacts directly. This entity allows site owners to directly collect customer analytics data, remember language settings, and carry out other useful functions that help provide a good user experience. There are a variety of ways to perform first-party tracking, for example, user accounts, first-party cookies or caches. In first-party cookie tracking, site owners record user information such as username, passwords, and items added to the cart by attaching a unique string to the user browser. For example, Google tracks user interests via the search engine. When a user enters a query in the search bar, Google keeps a record of this entry through login credentials or information such as IP addresses, caches, or cookies. It then

shows related links and advertisements in subsequent searches. A similar method is adopted by social networks such as Facebook, where a user could be tracked via interests shown through likes, comments, or posts. The user can control this tracking through security and privacy settings offered by these service providers.

Akin to websites, mobile app developers can also be considered first-party tracking entities as they have an ability to capture device data (e.g., sensors) and user information or activities through their apps. To capture device data, developers make API calls to multiple sensors like microphones, cameras, GPS, accelerometers, and touch. Similarly, device information such as phone numbers, current location, or unique phone ID number can also be extracted through APIs. However, in most cases, users' consent is required before extracting such information. The consent can be acquired by displaying app permissions and policies and getting explicit acknowledgment from a user.

First parties have several potential incentives to perform tracking. For instance, a first party wants to personalize user experience across sessions, detect frauds, or conform with law enforcement requiring websites to log user activities for fraud prevention and anti-laundering. However, there are cases where first-party websites voluntarily sell user identities. For example, Datalogix buys user information from companies, compile user dossiers, and then use it to target advertising [32]. Sometimes, a first party can also act as a third party (discussed in Sect. 9.3.2.2); for instance, logging in to a website using a third-party service such as Facebook or Google allows the website to request your data from them.

9.3.2.2 Third-Party Tracking

Third-party tracking is performed by the entities that track users across different services, for example, websites. It can also be an entity that provides resources while a page is being displayed. Typical resources are the content embedded in the page or external content accessed by a script running on the page. Third-party tracking offers several benefits to service providers, such as better audience targeting, boosting company recognition and reputation, or increasing its ROI. For instance, Google Analytics is a third-party entity used by more the half of the websites to gain aggregated statistics such as the business's performance, user experiences, user activities, and traffic records. This means that during any given browsing period, it is likely that at least some of that user's activity is being tracked by website owners, which is sent to Google Analytics for further processing. The processed data is then returned to the website owners to provide insights into their website traffic (e.g., geographic region and what type of device is being used) and user activity (e.g., page views and link clicks). Hence, aggregating data from various sources (e.g., websites, surveys, or publicly available information) can provide rich information in both breadth and depth. This allows a service provider to grow their targeted audience's size by including new prospects (e.g., who purchase similar or complementary products or services from a direct competitor or partner company).

Third parties have a range of motivations, which can be grouped into six main reasons, mentioned below:

- **Advertising** is one of the most common reasons to track and identify users online. In order to sell products, gain revenues, or increase product awareness, businesses and companies build associations with ad networks. However, it is essential to profile users and target the right ads on a website to be successful. For example, a user interested in buying a pair of shoes of a specific brand should be shown ads related to that brand.
- **Third-party measurement and analytics services** allow first-party websites to better understand their users by getting statistical information on demographics, content view distribution, and more. Third-party measurement services provide such analysis either using a paid or free analytics model. In a paid model, an analytics company takes precautions to silo data between clients, whereas in a free analytics model (e.g., Google Analytics), aggregated traffic statistics are sent to service providers to improve their content or enhance their services.
- **User engagement** can be increased via social networks, which allow service providers to offer personalized content and single sign-on services to their customers. These services either use cookies or require users to log in to their social network accounts and thus inevitably track and identify users. Examples include Facebook's like and comment widget and Google's like button. These features are offered for free to increase user engagement and to conduct market research. Moreover, there are social services that exist almost exclusively in a third-party context [33]. For instance, Disqus is a worldwide blog comment hosting service that offers features, such as social integration, social networking, user profiles, spam and moderation tools, analytics, and email notifications, to websites.
- Third-parties offer **Customized Content** such as video, maps, news, weather, stocks, and other media for embedding into websites. YouTube, for example, offers third-party widgets to generate revenue through in-widget advertising. Many others, such as the Associated Press, also charge for their content.
- **Content distribution** is yet another motivation for a third party to track and identify users. Content distribution networks, such as Akamai, help service providers distribute customized content to users based on their interests and profile.

There are potentially more intricate privacy issues with third-party tracking than with first-party tracking. For instance, users sometimes provide personal information such as contact details, email addresses, and billing information to a first party, which is sent to third parties for processing and detailed analysis. Hence, working across first-party providers, this third party then also has the ability to identify users across multiple website domains, thus providing much information about users, something they may be neither aware of nor comfortable with.

9.3.3 Tracking Techniques

In recent years, online tracking techniques have been extensively studied in academia. In contrast, only a few of them have been deployed online. We classify these tracking techniques into two categories: *(i) deployed tracking techniques* and *(ii) potential tracking techniques*. The deployed tracking techniques are widely used for online tracking and have been employed at a large scale in the mobile and web industries. On the other hand, potential tracking refers to the (advanced) mechanisms proposed by researchers in academia in an attempt to identify privacy leakages in mobile and web platforms.

9.3.3.1 Deployed Tracking Techniques

Deployed tracking techniques use IP addresses, cookies, Javascript, cache, and more for user identification purposes. In general, these tracking techniques operate as follows:

- **Cookies** are texts stored by a user's web browser and transmitted as part of an HTTP request. Cookies are essential to managing long user sessions, and they can be used to identify a user's browser uniquely. Service providers can use cookies to collect users' web activity. An example is Analytics Cookies (**_utma, _ga, _utmb**) that identify users or sessions and are used by website publishers to understand how people are using their website. Another particular form of cookie is a *persistent cookie*, which stores identifying information, such as user preferences, for an extended period. Similarly, *third-party cookies* are set while fetching website content, such as images, frames, and Javascript. *Cookie syncing* is another type of cookie where unique identifiers are correlated to identify a user in an external database for purposes discussed above. All these types of cookies are distributed and retrieved across multiple website domains allowing companies to build detailed profiles of users' interests, for example, spending history or frequently visited places such as restaurants. Intimate knowledge of users' personal preferences and private activities might eventually be used to brand them as members of a particular group, which could have serious privacy implications.
- **Javascript** codes can be loaded both from first- and third-party domains. They are widely used by ad networks, content distribution networks (CDNs), tracking services, analytics platforms, and online social networks [34]. They can track information about browsers such as cached objects, history of visited links, user-agent strings, or language preferences. In addition, they can read from and write to a cookie database or reconstruct user identifiers. Such information enables servers and third-party domains to track users using HTTP requests regularly. The dynamic nature of Javascript also allows service providers to construct a behavioral profile of a user. For example, through Javascript event handlers, it is

possible to obtain information about a user's mouse clicks and movements, and scrolls [35].

- **Caching** stores the content of webpages and other information in the browser to minimize latency and redundant network activity. This technique improves performance, as it becomes possible for a server to associate a unique tracking identifier with each client requesting content for the first time. A server can then use Javascript and standardized messages to check if the content is cached or not, to identify a user. This technique can be implemented for resources like images or fonts and is difficult to avoid unless the cached content is regularly cleared, for example, when closing the browser. For instance, Acar et al. [36] showed that 146 websites from Alexa's top 10,000 websites track users through fonts. Google was one of the domains that used fonts to track users to ensure quality and improve Google products and services [37].
- **Supercookies** also known as *unique identifier headers* inject user information into packets, which are then sent from a user device to a server. Some prominent supercookie types are *Flash Cookie* and *EverCookies*, where the former is maintained by the Adobe Flash plugin, and the latter is a combination of various tracking mechanisms. *Local Shared Objects (LSOs)* are supported by browser plugins, which can track users using unique identifiers. These objects are invisible to the browser, and therefore, it is impossible to examine their content. LSOs are retained in the browser even when the user deletes cookies and browser storage. For this reason, LSOs are used to store copies of browser cookies or other unique identifiers. All these types of supercookies contain unique identifiers allowing trackers to link records in their data to track browsing history and browsing behavior (e.g., visited websites including the length of stay). In 2014, Verizon and AT&T were found to be quietly tracking the Internet activity of more than 100 million cellular customers with "supercookies," which allowed the companies to monitor which sites their customers visit, cataloging their tastes and interests [38]. In other words, network providers (who are supposed to provide a content-agnostic service) were inspecting the contents of users' Internet traffic without their knowledge. Such tracking aimed to facilitate advertisers to display ads based on individual Internet behavior, however, considered against the GDPR.
- **Stateless tracking** allows websites to track users based on information such as user agent, fonts, screen resolution, and more. Standard techniques for stateless tracking are as follows: (i) canvas fingerprinting detects minor differences in display hardware by reading back rendered text from a storage area mapped to the display, (ii) font/plugin fingerprinting involves detection of fonts or plugins supported by a browser, (iii) MediaStream Fingerprinting is performed through Media Capture and Streams API that generates a unique stream identifier, (iv) WebRTC determines local IP address behind any firewall and can generate a unique tracking identifier, and (v) user agents/IP address in combination can be used to identify the user behind a browser. Although some of these techniques

individually produce medium-entropy identifiers, it has been shown that a combination of these is unique enough to generate a high-entropy identifier.[2]

We refer interested readers to [32], for a survey and in-depth study of online tracking mechanisms.

9.3.3.2 Potential Tracking Techniques

The abovementioned deployed tracking techniques have been extended by researchers either by using additional identifiers or by using advanced classification technologies, such as machine learning. The pioneering work in the threat of tracking dates back to Sweeney, who showed for the first time that coarse-grained information such as birthday, gender, and ZIP code could uniquely identify a person [39]. This work was followed by several studies that provided measurement insights into web and device tracking. The success of such methods is a clear indication that anonymization techniques to protect the privacy of individuals may fail if the collected data contains unique combinations of attributes relating to specific individuals. This section presents the online tracking technologies proposed by researchers and categorizes them based on the tracking medium: web browser, mobile phones, or other devices.

Web Browser-Based Tracking Techniques
In the past decade, several studies measured and analyzed web tracking. The authors of [40] provided an early insight into web tracking, followed by a continual increase in third-party tracking techniques. Also, [41] quantified the uniqueness of web browsers based on user agent and/or the browser configuration (plugins, fonts, cookies, screen resolution) and showed that 90% of browsers could be uniquely identified by the user agent, cookies, time zone, plugins, and fonts. The algorithm was able to detect returning browsers, even if some features changed over time.

Following this, [42] quantified the amount of information revealed by host identifiers, including IP addresses, cookies, and user login IDs. Authors used month-long datasets of a web-mail service and a search engine for the analyses. Further, they discussed the implications of cookie-churn on privacy and security, along with the utilization of host fingerprinting for improving security. An extended approach presented in [43] showed that cross-browser fingerprinting could achieve high uniqueness if the operating system collected enough data.

The authors of [44] performed a large-scale analysis of web browsing histories to track users. They were able to detect 97% of browsers by inspecting only four web pages in the browser history. Akin to this, [45] explored browser fingerprints

[2] Medium-entropy identifiers refer to the attributes/features that give limited information about a device or a user, that is, low information gain to the trackers. On the other hand, high-entropy identifiers refer to the attributes/features that contain rich information about users or devices, that is, high information gain.

validity by collecting more than 100K fingerprints composed of 17 attributes. Their results showed that HTML5 and Canvas API offer highly distinguishable features. A fingerprint technique based on the measurement of on-screen dimensions of font glyphs is proposed in [46].

A crawler-based measurement study of online tracking at 1M websites was reported in [47]. The analysis was based on stateful (cookies) and stateless (fingerprinting) tracking, the effect of browser privacy tools, and data exchange between different sites (cookie syncing). The authors developed an open-source privacy measurement tool, which simplifies data collection for privacy studies on a scale of millions of websites. Similarly, [48] studied the effect of third-party HTTP requests on the top 1M websites and showed that Google could track across 80% of websites through third-party domains. It has been shown that 80–90% of browsers can be uniquely identified. Besides HTTP cookies, other entities such as Flash cookies, WebGL, and HTML5 were also used as a tracking medium [22, 49].

It is important to mention that several side-channel and timing attacks have been launched on web browsers to leak the browser histories and cache information [50, 51]. These attacks can de-anonymize users in social networks, uncover user data, or reveal data to service providers or ad networks. Two different studies, [52] and [53], showed that usernames and online social profiles could uniquely identify user profiles and link users across different social platforms. In these works, fingerprinting was based on device configuration, device settings, and device hardware.

We summarize popular web-based tracking techniques as follows.

- **Web Tracking Measurement Studies** crawled data using Firefox extension and plugins [22, 40, 54] or open-source tools such as Open WPM [47, 55] and webXray [48, 56]. These mechanisms crawled attributes mainly including first- and third-party cookies, JavaScripts, canvas font, audio, JSON, PHP and CGI scripts, tracker-owned cookies, site-owned cookies, and HTML5 Local storage.
- **Web Browser Fingerprinting Techniques** used information gain, entropy, and k-anonymity to fingerprint the browsers [41, 45, 46, 49]. The attributes which contributed toward fingerprints mainly include user agent, cookies, timezone, screen resolution, MIME types, system fonts, WebGL, and HTTP headers.
- **Cross Browser Web Fingerprinting** used anonymity sets, entropy, and correlation as fingerprinting metrics [43]. The features used are user agent, OS, screen resolution, basic fonts, and timestamp.
- **Web-Based Device Fingerprinting** used host tracking graphs, entropy, and battery-reading techniques for fingerprinting [42, 44]. These techniques used attributes such as user agent, IP address, browser cookies, battery level, readouts, and charge/discharge time.
- **Online User Profiling** used information surprisal, entropy, and Markov chain as fingerprinting mechanisms [52, 53]. The information used for profiling includes gender, age, usernames, city and status.

Mobile-Based Tracking Techniques

Mobile device fingerprinting is a recent technique used by companies to profile device data or user interests. In general, the techniques mentioned above for browser fingerprinting can also be used for mobile tracking. However, studies revealed that mobile browsers do not have such distinguishable features as plugins and fonts; thus, requiring fingerprinting methods that are specifically designed for mobile devices or browsers [41]. Thus, several studies proposed alternative methods to fingerprint mobile devices. These techniques utilize different physical characteristics of a mobile device, for example, camera, sensors, microphones, and speakers. For instance, a study in [28] used the vibration motor to develop accelerometer fingerprints and then applied machine learning to extract the frequency and time-domain features. These features were able to distinguish mobile devices with 99% of accuracy.

Authors in [57] proposed a fingerprint mechanism to uniquely identify smartphones based on motion sensors (accelerometer and gyroscope) and inaudible audio stimulation, along with a mechanism to obfuscate the fingerprints by calibrating sensors. Noise-based sensor fingerprinting for mobile devices has also been discussed in [58–60], which focused on acoustic components such as speakers, microphones, or cameras. These techniques require access to the microphone, which needs separate permission. Authors in [61] utilized the noisy nature of hardware sensors such as accelerometer and microphones. Similarly, images taken by a mobile phone camera can derive a noise pattern that is considered to be different in each device sensor [62, 63].

A study conducted in [64] focused on mobile users' identification and tracking based on touch-based gestures. Their fingerprinting mechanism extracted statistical features from swipe, keystrokes, taps, and handwriting gestures and showed a true positive rate of 93% to detect returning users. Some studies have also focused on privacy-preserving online behavioral targeting for various purposes, including advertising, spamming, and political interests [65, 66]. Another work [67] analyzed 59 mobile device fingerprints and concluded that "the fingerprints taken from mobile devices are far from unique and targeting." However, they did not consider the canvas test for fingerprinting. Authors in [68] presented a new side-channel attack against smartphone keyboards that support gesture typing. They identified returning users with 97% accuracy using a set of 35 sentences, and the system also correctly predicted sentences.

A number of studies have focused on identifying mobile user traits and characteristics using the information provided by mobile SDKs to third-party apps, such as the running apps, device model, and operating system [31]. A study in [69] showed that mobile devices can be tracked through personalized configurations (e.g., installed apps, top 50 songs, device, WiFi name) without involving hardware identifiers such as Unique Device Identifier (UDID), International Mobile Station Equipment Identity (IMEI), and others. A work in [70] showed the existence of a diverse set of mobile users using clustering and feature ranking. Their results

identified 382 categories of users based on their app usage patterns. We summarize popular mobile tracking techniques below:

- **Mobile Tracking based on Motion Sensors** used techniques such as bagged decision trees, machine learning classifiers, Gaussian mixture models, and k-nn classifier with common features such as spectral centroid, spectral skewness, spectral flatness, and average deviation [28, 57, 59, 71–73].
- **Mobile Tracking based on Audio Sensors** used Euclidean distance and L2 distance with main features such as sensitivity parameters, vector aptitude, feedback ratio of different frequencies and harmonics [60, 61].
- **Mobile Tracking based on Camera Sensors** used SVM, Photo-Response Non-Uniformity (PRNU), and Pearson correlation mechanisms [62, 63]. The techniques used 81 features (i.e., 3 RGB channel * 3 wavelet components * 9 central moments).
- **Mobile Tracking based on Touch Sensors** used cosine similarity, entropy, information gain, and recurrent neural network mechanism [64, 68]. Authors used 50 extracted touch features such as x-coordinate, y-coordinate, finger pressure, and finger area.
- **Mobile Tracking based on Mobile Browsers** used Open AM algorithm with features such as screen dimensions, color depth, installed plugins, user agents, and timezones [67].
- **Mobile Tracking based on Personalized Configurations** used Jaccard similarity coefficient, k-means clustering, entropy, and SVM as main mechanisms [31, 69, 70]. Features mainly include device model, device ID, username, installed apps, etc.

Device Tracking Based on Network Properties: Some of the fingerprinting techniques have also used properties, such as network configuration or traffic records, for a device or host tracking. One of the prominent works on remote device fingerprinting was presented in [74] that proposed a method to measure device clock skew using ICMP and TCP traffic. Some works also deal with wireless traffic; for example, radiometric analysis of 802.11 transmitters [75], signal phase identification of Bluetooth transmitters [76], or timing analysis of 802.11 probe request frames [25]. For example, [75] utilized manufacturing defects in hardware to identify the device and, by association, the end-user. Many efforts on tracking wireless devices focused on other hardware characteristics, such as radio frequency and drivers [77, 78]. While these techniques can also be used to identify smartphones, these calculations are also resource intensive and require user cooperation. In addition, identifiers such as network names and IP addresses also help in host fingerprinting [75].

9.3.4 Behavioral Tracking: State of the Art

Behavioral-based tracking refers to constructing user profiles and uniquely identifying the users through their gestures to perform certain activities [79]. Such gestures are collected via many modalities such as touch, motion, GPS, camera, mouse, search queries, writing pattern, and more. Examples of such information include the location of a user at a particular time, user-touchscreen interaction, duration of the calls, and dialed numbers. Such profiling could be used by data custodians, receivers, or consumers, in order to provide personalized services to their customers with the goal of increasing revenues. For instance, advertising companies use user behavior profiles, user interests, characteristics, or activities to display relevant advertisements to the user [80].

The ability to distinguish behavioral biometrics is a new form of tracking. Behavior-based tracking has the ability to continuously and surreptitiously track users while they are interacting with their devices. As opposed to "regular" tracking mechanisms based on cookies, browser fingerprints, logins, and similar, which track virtual identities or browser profiles, this type of tracking is subtle. First, while regular tracking deals with virtual identities and online profiles, behavior-based tracking has the potential to track and identify the actual (physical) person operating the device. It can track multiple users accessing the same device by profiling user behavioral activities such as touch gestures [81]. Second, behavior-based tracking has the ability to track users continuously. Third, it also leads to cross-device tracking, where the same user can be tracked on multiple devices, and user data can be collated and used to build more encompassing user profiles. However, implementing such tracking requires a more generalized approach, requiring, for example, to validate the stability of features across devices [64]. On the other hand, the ubiquity of smart devices and the fact that any web service can extract data from touch and motion sensors make behavioral-based tracking quite achievable. This not only represents a valuable source of information for analytics and ad services but also for app developers who can use the information to track individuals on a single device or across devices. Table 9.1 summarizes the distinction between behavioral tracking and other tracking methods.

Nevertheless, behavioral-based tracking is equally beneficial to users and service providers. Some argue the benefits of behavior-based tracking as a way of receiving useful information, for example, relevant ads or health monitoring. For instance, monitoring a phone's motion might reveal changes in gait, which could be indicators of ailments or depression [82]. Another benefit of behavioral tracking is continuous or implicit authentication of users on mobile and web platforms. Implicit authentication is a mechanism to continuously authenticate users while they perform activities on mobile or web platforms. This type of authentication continuously monitors users through behavioral biometrics such as touch swipes, taps, keystrokes, or stylometric features, to verify a user's legitimacy with high accuracy. The usability and deployability of implicit authentication schemes without compromising security have made them an attractive alternative to legacy password systems [82].

Table 9.1 Behavioral-based tracking versus other tracking methods

Behavioral-based tracking	Other tracking methods
Aims to identify or track user or his activities	Aims to identify devices or browsers
Constructs user profiles based on their gestures to perform certain activities	Constructs profiles based on device or browser configuration, specifications, or settings
Has a potential to track the physical identity of users	Has a potential to identify virtual identities of users
Continuously track users through their behavioral actions	Track device or browser only when a certain action is performed, for example, when a website is visited
Is ideal for cross-device or cross-system tracking	Is less suitable for cross-device or cross-system tracking

9.4 Personalization via Online Tracking

As mentioned earlier, Internet users are increasingly being tracked, and their personal data are extensively used in exchange for services. In the current era, when people use real identities to communicate on the Web, maintaining privacy has become a complicated challenge. Service providers are using a variety of personal information to personalize their content and services. The privacy challenge becomes more critical with the dissemination of smart phones and devices offering new possibilities for personalization. On the other hand, personalization algorithms and technologies are steadily improving, making behavioral profiling more powerful, yet raising a multitude of privacy challenges.

To understand the personalization system, Fig. 9.2 shows an exemplary working diagram of an advertisement network system. There are three main entities in an advertisement network system: the publisher, the advertiser, and the ad network. The publisher is an entity that owns a website or service; the advertiser is an entity that wants to advertise to users; and the ad network collects advertisements from an advertiser, displays them on a publisher's website, and connects advertisers to users with relevant demographics. If a user clicks on an advertisement, the ad network collects money from an advertiser and pays part of it to the publisher. It is thus important for the ad network to generate accurate and complete profiles of users, in order to increase the click chances and maximize the revenues. These three entities also exist in a mobile environment, where a mobile app developer acts as a publisher, while the roles of an advertiser and ad networks remain unchanged.

It should also be noted that, while the above example is tailored to the advertisement context, it is similarly applicable to other applications. Consider other scenarios, such as multimedia content consumption on YouTube or Spotify, a news dissemination platform, or even a student eLearning environment. In all of them, the three entities—content provider, content consumer, and the intermediate network—can be easily identified and the need for accurate user profiles for an enjoyable and engaging service is evident.

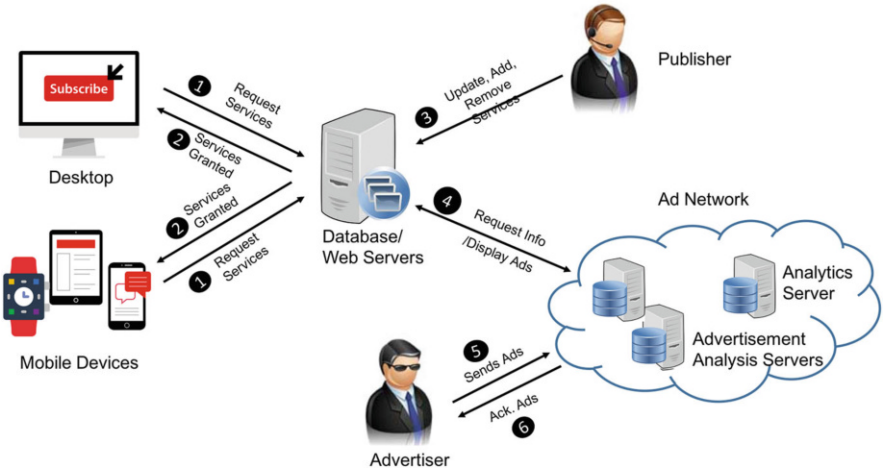

Fig. 9.2 Ecosystem of advertisement network

9.4.1 Relationship

Personalization is hard to achieve without losing privacy since a service provider needs users' personal information to tailor or customize services. Research has shown that users are willing to share their personal interests or information in exchange for the apparent benefits of using personalized products or services [83, 84]. To build trust, some service providers promise to ensure the anonymity of their customers for the usage of their services, and in some cases, the anonymity is guaranteed for a lifetime. On the contrary, research shows that it is difficult to guarantee anonymity as linking anonymized data to other databases with personally identifiable information leads to the (re)identification of a user [85]. Therefore, privacy risks are not just limited to a particular service provider, rather these risks are pervasive concerns where personal information provided by users to different services could be linked together to track/identify them ubiquitously. Authors of [14] discussed the risks associated with recommender systems. The authors argue that privacy breaches are either due to direct data access or due to data sharing with third parties. In both cases, the effects of privacy breaches can be significant, such as exposure of sensitive information, reidentification of anonymized data, leaks through the shared device, or service inference by the recommender.

In [86], authors link privacy to three different personalization categories: social, behavior, and mobile.

- In a **social-based personalization**, providing privacy is a major concern because of three main reasons: (i) users are willing to reveal more information, (ii) social networks compromise not only a single user's privacy but also their friends' privacy, and (iii) social networks can reveal potentially embarrassing

information. There have been several cases where an employee's misuse of social media has led to their dismissal. For example, various incidents resulted in employee termination from firms based on their post or comments on social media [87, 88]. According to one survey, 17% of companies with 1000 or more employees report issues with employees' use of social media, whereas 8% of those companies fired employees because of information released on social networks [89].

- **Behavior-based personalization**, where information about observable user activity is longitudinally collected and harnessed for personalization purposes, poses several privacy risks. These include unsolicited marketing, personal information being shared with third-party providers without users' consent or knowledge, and in some cases, being inadvertently revealed to other users of the same device. For instance, users who share a computer and or a Web browser may view each other's ads if cookies are used by websites to identify users. Another risk involves linking behavioral profiles to server-side user accounts so that advertisers can target users across different devices.
- **Mobile-based personalization** has increased with the spread of smartphones and phone sensors. With this, the ability of service providers to continuously track users has also grown. Sensor data has been used in various ways for personalization. One way is the improvement of search results, such that search results displayed on a smartphone are tuned according to the user's location, highlighting nearby venues and services. Similarly, the installation of various apps on mobile phones conveys user interests, helping app developers to show targeted ads. Authors in [90] performed a measurement study of in-app advertisement and showed that GoogleAdMob has a higher proportion of targeted than generic ads. Privacy leakages in mobile-based personalization are more significant, mainly because mobile devices are carried around all the time and are increasingly being used for sensitive operations like personal communications, dating, and banking. Therefore, privacy concerns regarding what information is collected for ad personalization are rather serious.

9.4.2 Privacy Implications of Personalization

Although personalization via online tracking has been performed for a number of reasons that bring tremendous value, it also raises serious privacy concerns having subtle and far-reaching consequences. Researchers, civil organizations, and policymakers have identified several ways tracking can cause privacy leaks.

Global surveillance, performed by the government for security reasons or by companies for commercial benefits, is one such privacy risk. Between January and June 2014, the US government made 12,539 requests for 21,576 persons' information from Google, including search history, and Google complied with 84% of them [91]. According to the internal National Security Agency (NSA) presentations [92], the American NSA and British GCHQ use cookies (one of them

being Google PREFIDs) to investigate the online activity of users. The government agents are first granted access to Internet links, and then they use cookies to differentiate flows generated by different users within the same Internet connection. These cookies help them track user locations and denounce users who have unauthorized access to the network. The presentation also revealed that some NSA divisions engage with private companies and Internet service providers to collect data, which are later used for hacking into people. Another presentation revealed that the NSA uses DoubleClick cookies to identify TOR users [92]. A program named HAPPYFOOT by the NSA was designed to map users' Internet addresses to their physical locations. By capturing the Internet traffic, the NSA gathers almost five billion records a day on the locations of cellphones around the world. That also allows the NSA to track how particular people travel and gain knowledge about their mutual relations by revealing co-travelers [93]. Such surveillance is a threat to privacy, but there may be chances that collected information is distorted and leads to incorrect decisions. The potential dangers could be an error, abuse, and lack of transparency and accountability [32].

Unwanted profiling, performed by service providers to personalize content for users, is another risk. A news site may display news matching users' previous items, a merchant may propose products based on users' previous shopping, or a search engine may refine results based on users' previous queries. Often, such profiling may seriously impact users. For example, it was shown that a person discovered his teenage daughter was pregnant when she received advertisements for baby food. The teenager was profiled as pregnant based on her shopping behavior [94]. Similarly, Gmail was shown to use words from the sent and received emails to target ads. The emails were scanned without a user's explicit permission and used to identify the themes and trends for ad targeting [32]. The Facebook Beacon advertising program faced a federal class-action lawsuit because users were automatically opted into having purchases disclosed to friends and networks [95].

Reidentification of anonymized public data is required in several business applications and research studies to improve the provided services by utilizing the available information and rich user data. However, studies have shown that users could be identified even from anonymized datasets through inference analysis by an eavesdropper. A few examples involving such threats are the reidentification of users in the anonymized AOL search histories, Netflix training data that was attacked, and Massachusetts hospital discharge data [96–98]. For instance, in an open competition for the best collaborative filtering algorithm in 2009, Netflix disclosed data records of 480,000 customers "anonymously" in an attempt to create a smarter recommendation algorithm. The data contained subscribers' information, including gender, zip code, age, unique subscriber ID, the movie title, year of release, and the date on which the subscriber rated it. Despite being anonymous, researchers were able to reidentify sensitive information about people, as in the case of a closeted lesbian mother who sued Netflix for disclosing her sexuality to the public through rented movies such as Brokeback Mountain or Passion of the Christ. Similarly, when AOL released anonymized search queries of its customers, a 62-year-old widow was identified living in LilBurn GA, Georgia, United States. The

lady frequently searched for her friend's medical ailment and loved her three dogs. These examples show that it is possible to violate the users' privacy by tracking their activities, thereby inferring their personal profiles. Thus, users' privacy is at risk when their data can be distinguished from other users and linked with high confidence based on the user's previous history.

Personalized search, which offers the benefit of presenting information that the user wants to see based on their queries, is another reason to track. However, it has been shown in [85] that even anonymized search queries could lead to the identification of users and their interests. The ability for a search company to efficiently track and record users' search habits and tie them directly to their identity has profound privacy implications. For instance, search engines may know the current situation of a user (e.g., illness, depression, studying, startup business, or looking for jobs) through their searches and show them results related to their situation (e.g., recruitment websites, training workshops, discounted medicine prices).

Lastly, tracking was found to be the reason for **price discrimination** based on geographical location, affluence of the user, and the referrer. Examples include credit card interest rates, hotel bookings, and insurance coverage. In [32], authors provided a detailed overview of how such implications occur. For instance, Capital One Financial Corporation differentiates car loans' interests based on the browser used by the prospective customer (3.5% for Firefox, 2.7% for Safari, 2.3% for Chrome, and 3.1% for Opera). Similarly, Orbitz Worldwide Inc. differently sorts out the hotel advertisements depending on the type of computer used by the customer. Orbitz found that Mac users tend to spend around 30% more on hotel bookings than PC users. Using this fact, more expensive hotels are advertised to Mac users, while the cheaper ones to the PC users.

9.4.3 Balancing Privacy and Personalization

It is reasonable to expect that users would be more inclined to share their data with service providers and use personalized services if the user information is collected and treated fairly. However, striking a balance between privacy and personalization is quite a challenge. Researchers, businesses, and nonprofit organizations have made a continuous effort to provide efficient solutions to overcome user privacy/tracking issues. Some of these efforts have resulted in privacy design principles, privacy tools, and features. In this section, we discuss the technological measures that could be taken to minimize tracking via personalization.

9.4.3.1 Privacy-by-Design

Privacy-by-Design is deemed an essential step toward better privacy protection. It is based on the idea that privacy requirements should be taken into account while

designing a system. As with any process, privacy by design should have well-defined objectives, methodologies, and evaluation metrics.

Consent-Based Mechanisms Consent-based mechanisms are one way to obtain privacy-by-design. These mechanisms inform and obtain users' consent before collecting and processing the data. According to General Data Protection Regulation (GDPR), a consent is "any freely given, specific, informed and unambiguous indication of the data subject's wishes by which he or she, by a statement or by a clear affirmative action, signifies agreement to the processing of personal data relating to him or her" [21]. Therefore, it is necessary for a user to know which data is collected and for what purpose. The most widespread mechanism for user consent on the web is probably the cookie header banner, which is displayed on all web pages and invites user to make a choice of accepting or refusing cookies.

Another widespread mechanism is to take consent through browser settings which offer four options, that is, accept all cookies of a websites, accept cookies set or accessed by first party, accept cookies set by first party only, or accept no cookies. A "tag manager" is also a technical implementation of the cookie consent that could block third-party scripts if consent has not been obtained. One key issue with consent-based mechanisms is that the entity that informs users is often not the only entity to track users. For instance, third-party trackers also collect and share information about users, which the first party may be not be aware of. In this situation, some methods that are less often employed are first- and third-party consent tools, which are used to make an agreement between parties, explicitly stating what user data will be obtained and for what purpose.

Obfuscation Methods Several obfuscation methods have been proposed as the means to maintain user privacy in recommender systems. These mainly include distribution, aggregation, anonymization, identity management systems, privacy proxies, encryption mechanisms, and differential privacy. One strategy is to distribute user data across a set of machines; however, this solution aggravates personalization based on data of other users [99]. Another strategy is to use the encrypted aggregation of user data [100, 101]. Privacy-preserving approaches like differential privacy and k-anonymity are the widely used privacy-preserving solutions. Differential privacy mathematically guarantees that anyone seeing the result of a differentially private analysis will essentially make the same inference about any individual's private information, whether or not that individual's private information is included in the input to the analysis [102]. It provides a mathematically provable guarantee of privacy protection against a wide range of privacy attacks mainly including differencing attacks, linkage attacks, and reconstruction attacks [103]. Similar to DP, k-anonymity also guarantees privacy by holding a property that a released dataset is k-anonymous if the information for each person contained in the dataset cannot be distinguished from at least $k - 1$ individuals whose information also appear in the dataset [104].

A study in [105] investigated the effectiveness of different obfuscation strategies and policies for online social networks and proposed a novel obfuscation strategy

that does not require knowledge about the adversary classifier. Authors of [106] and [107] proposed methodologies that prevent inference attacks by distorting data before making it publicly available. In [108], authors proposed a utility-aware obfuscation framework that limits the risk of disclosing sensitive information from sensors data. Similarly, work proposed in [109, 110], and [111] tried to protect user location data by generating fake privacy-preserved location traces. In another recent paper [112], author proposed an obfuscation scheme [27] to defeat fingerprinting based on motion sensors.

Pseudonymous personalization is a basic yet common approach to hide true user identity. It allows people to use the same pseudonym across different sessions and to create or maintain more than one pseudonym. This helps users separate different aspects of their online activity and control which service provider can access their persona [113, 114]. However, anonymity is difficult to maintain when payments or nonelectronic services are involved. It has also been shown that hiding explicit identities like usernames and emails are not sufficient to prevent tracking. There are cases where users have been identified through their anonymized data, hence revealing personal/sensitive information about them [85].

Client-side personalization is another way to prevent online tracking. This type of privacy preservation implies data storage and subsequent personalization processes to take place on the client-side [115]. Since data collection and processing occur at the client side rather than the server-side, users may perceive more control over their data and lower privacy risks. However, the challenge with this approach is that existing personalization algorithms need to be redesigned to fit the client-side model [12].

User controls and feedback is another way to preserve privacy in personalized systems. Studies conducted in [116, 117] suggested adding scrutability to user modeling and personalized systems. The term "scrutability" signifies the users' ability to understand and control what goes into their user models, what parts from their models are available to various services, and how the model is managed and maintained. This allows users to restrict service providers from accessing their sensitive data. However, achieving such a level of balance is currently challenging due to poor user understanding of these notions.

9.4.3.2 Privacy Tools

A number of browser tools and plugins have been developed to protect users from tracking. These tools perform various functionalities such as detecting or blocking lists of third-party trackers, informing users how much information is revealed to trackers, allowing only executable content from trusted domains to run, detecting flash cookies and deleting them, and more. ENISA provided a detailed analysis of online privacy tools in [118]. The study analyzed several web portals that are listing and/or recommending the use of specific online privacy tools (e.g., for secure messaging, anti-tracking, and encryption). There is also a Tracking Protection List (TPL) approach that contains addresses of misbehaving tracking sites published by

various organizations. Other ways to protect information include tools like private browsing modes of major browsers and anonymity networks.

Do Not Track (DNT) Major browsers implement the DNT (Do Not Track) methodology to show websites that they are forbidden from tracking. DNT is a technology and policy proposal that enables users to opt out of tracking by all third-party websites they do not visit, including analytics services, advertising networks, and social platforms [119]. Technically, the implementation of DNT is simple; a browser sends a DNT header in every HTTP request to websites the users wish to opt out of tracking. This includes web pages and all the objects/scripts embedded within a page. However, it is up to the discretion of an advertiser to respect user preferences.

Similar to DNT, some other tools have also been proposed to anonymize web search queries. For example, TrackMeNot (TMN) [120] is proposed as a Firefox plugin to randomly issue dummy queries from predefined Rich Site Summary (RSS) feeds. GooPIR is a standalone application for noise addition to Google queries [121], which modifies the user queries by adding dummy keywords, and then the search results are re-ranked locally based on the original user queries. PRivAcy model for the Web (PRAW) [122] is another technique, which continuously generates fake queries in different topics of interest of the user. This is done by generating user profiles from user queries and corresponding responses and thus the fake queries added will be in the general area of interest of the user to make the distinction between real and fake queries difficult.

Decentralized Ad Platforms A few behavioral advertising systems, like Adnostic, PrivAd, and RePriv, consider privacy as a design requirement. The main objective of these systems is to limit tracking, while still serving behavioral advertisements. For instance, PrivAd preserves privacy by maintaining user profiles on the user's device, thus minimizing the information released to the ad network. A trusted third party anonymizes the network addresses of clients whereas encryption prevents the proxy from viewing client messages. As such, PrivAd offers privacy against profiling, ad dissemination, auctions, click fraud, view and click reporting, and click anonymization [65]. Similarly, in Adnostic, the browser continuously updates user profiles [66], allowing the ad network to offer several ads to the browser, where the browser picks the ad most relevant to the profile. In addition, the principle of privacy-by- design has also been introduced by some web browsers such as Brave, which is a free and open-source web browser that aims to block ads and website trackers [123]. Brave also introduced the first advertising platform that puts the user in control with privacy by design and does not leak the user's personal data from their device. The ad matching happens directly on the user's device, such that the user's data is never sent to anyone.

User Agents User agents prevent tracking by providing users with relevant choices. Most user agents include functionalities that allow users to examine cookies associated with a domain or a web page, showing expiration date duration, their contents, and the associated host domain [21]. Such information can be presented

as user agent settings through a user interface to get valid consent from the user. This has already been implemented by a browser extension that uses the DNT Consent API to take consent from user before sending or receiving any data from the browser. Similarly, the Content-Security-Policy API (CSP) is another tracking prevention tool that prevents cross-site scripting, click-jacking, and other code injection attacks. CSP provides a standard method for first-party services to declare specific types of content that user agents should be allowed to load on that website—covered types are JavaScript, CSS, HTML frames, web workers, fonts, images, embeddable objects, etc. If any of these content types are provided in the source list within the CSP header, then user agent will load only that content type in a browser and block rest of the types. In this way, user agents can be told to block iframes from being loaded when they have not been explicitly allowed by the site designers or which refuse to respect the provided CSP. In general, user agents can prevent tracking at various granularity levels. This includes (1) items the user wants to block or take consent, like list of websites, tracking companies, (2) locations of blockage, (3) types of data, or (4) purpose of data.

Opt-Out Cookies Some tracking companies allow users to set opt-out cookies. If implemented properly, this option disables user tracking. However, opt-out cookies are not considered reliable, as they are not supported by all ad networks and are easy to interpret by those wishing to track users. Moreover, they have a limited lifetime, so they must be periodically renewed. These cookies are lost when the user cleans the cookies from their web browser.

Chapter 8 covers more details about privacy enhancing technologies, in a general sense. We recommend interested readers to go through the chapter for more information on privacy preserving solutions.

9.5 Conclusion

The ever-changing technological landscape, high user involvement, increased societal visibility, and amalgamation of services have made privacy difficult to maintain in a digital world. Preserving user identity from being tracked is a significant challenge nowadays and has become more complex with the advancement in technologies that have an ability to cross-link data sources to infer more information. Some examples aggravating the privacy concerns include location-based tracking, mobile sensors to identify location, behavioral features, interactions and gestures, and so on.

Moreover, the state-of-the-art data analysis methods and the exponentially growing computational resources available for data mining tasks are another potential obstacle for balancing privacy and personalization. For example, cloud-based data centers have the ability to process and compare user profiles among massive sets of records, to identify relevant information and make sense of it. As the user models and predictions become more accurate, and as the services increase their

reliance on these predictions, user privacy concerns may further increase. The propagation of online social network in our daily life also poses new challenges, as personalization processes are targeting not only online user activities but also the physical environment.

The proposed solutions to preserve privacy and prevent tracking have practical limitations that often preclude their developers from striking the balance between privacy and utility goals. Nevertheless, we would like to emphasize the need for technically encompassing, while also user-friendly, policy-compliant, and transparent, solutions. We believe that tracking-related privacy concerns will take a more prominent role and will attract research works and practical industry attention alike.

References

1. AP. 2018. Google has been tracking your movements even if you told it not to. https://www.news.com.au/technology/gadgets/mobile-phones/google-has-been-tracking-your-movements-even-if-you-told-it-not-to/news-story/bb9eb906387ffd2295e8b17b24b7d883
2. ur Rehman, I. 2019. Facebook-cambridge analytica data harvesting: What you need to know. *Library Philosophy and Practice*, 2497: 1–11.
3. Wong, J. C. 2019. Facebook discloses operations by russia and iran to meddle in 2020 election. https://www.theguardian.com/technology/2019/oct/21/facebook-us-2020-elections-foreign-interference-russia
4. BBC. 2018. Facebook's data-sharing deals exposed. https://www.bbc.com/news/technology-46618582
5. Hautala, L. 2019. These android apps have been tracking you, even when you say stop. https://www.cnet.com/news/these-android-apps-have-been-tracking-you-even-when-you-say-stop/
6. Zhang, B., N. Wang, and H. Jin. 2014. Privacy concerns in online recommender systems: Influences of control and user data input. In *10th Symposium On Usable Privacy and Security ({SOUPS} 2014)*, 159–173.
7. Berkovsky, S., and J. Freyne. 2015. Web personalization and recommender systems. In *Proceedings of the 21th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining*, 2307–2308.
8. Brusilovsky, P., A. Kobsa, and W. Nejdl. (Eds.) 2007. *The Adaptive Web, Methods and Strategies of Web Personalization*. *Lecture Notes in Computer Science*. Berlin: Springer.
9. Ning, X., C. Desrosiers, and G. Karypis. 2015. A comprehensive survey of neighborhood-based recommendation methods. In *Recommender Systems Handbook*, 37–76. Berlin: Springer.
10. De Gemmis, M., P. Lops, C. Musto, F. Narducci, and G. Semeraro. 2015. Semantics-aware content-based recommender systems. In *Recommender Systems Handbook*, 119–159. Berlin: Springer.
11. Berkovsky, S., T. Kuflik, and F. Ricci. 2008. Mediation of user models for enhanced personalization in recommender systems. *User Modeling and User-Adapted Interaction* 18 (3): 245–286.
12. Vallet, D., A. Friedman, and S. Berkovsky. 2014. Matrix factorization without user data retention. In *Advances in Knowledge Discovery and Data Mining - 18th Pacific-Asia Conference, PAKDD 2014, Tainan, May 13–16, 2014. Proceedings, Part I*, 569–580.
13. Erkin, Z., M. Beye, T. Veugen, and R. L. Lagendijk. 2012. *Privacy-Preserving Content-Based Recommender System*. New York: ACM.

14. Friedman, A., B. P. Knijnenburg, K. Vanhecke, L. Martens, and S. Berkovsky. 2015. Privacy aspects of recommender systems. In *Recommender Systems Handbook*, 649–688. Berlin: Springer.
15. Adomavicius, G., and A. Tuzhilin. 2011. Context-aware recommender systems. In *Recommender Systems Handbook*, 217–253. Berlin: Springer.
16. Lathia, N. 2015. The anatomy of mobile location-based recommender systems. In *Recommender Systems Handbook*, 493–510. Berlin: Springer.
17. Yang, W.-S., H.-C. Cheng, and J.-B. Dia. 2008. A location-aware recommender system for mobile shopping environments. *Expert Systems with Applications* 34 (1): 437–445.
18. Yu, J., J. Zhao, Y. Chen, and J. Yang. 2015. Sensing ambient light for user experience-oriented color scheme adaptation on smartphone displays. In *Proceedings of the 13th ACM Conference on Embedded Networked Sensor Systems*, 309–321.
19. Freyne, J., J. Yin, E. Brindal, G. A. Hendrie, S. Berkovsky, and M. Noakes. 2017. Push notifications in diet apps: Influencing engagement times and tasks. *International Journal of Human–Computer Interaction* 33 (10): 833–845.
20. Rogers, S., and P. Langley. 1998. Personalized driving route recommendations. In *Proceedings of the American Association of Artificial Intelligence Workshop on Recommender Systems*, 96–100.
21. Online tracking and user protection mechanisms. 2017. White paper, European Union Agency For Network and Information Security (ENISA).
22. Roesner, F., T. Kohno, and D. Wetherall. 2012. Detecting and defending against third-party tracking on the web. In *Proceedings of the 9th USENIX conference on Networked Systems Design and Implementation*, 12–12. Berkeley: USENIX Association.
23. European Commission (EU). 2016. Cookie sweep combined analysis report. https://ec.europa.eu/newsroom/article29/item-detail.cfm?item_id=640605
24. Rose, I., and M. Welsh. 2010. Mapping the urban wireless landscape with argos. In *Proceedings of the 8th ACM Conference on Embedded Networked Sensor Systems, SenSys'10, New York*, 323–336. New York: ACM.
25. Desmond, L. C. C., C. C. Yuan, T. C. Pheng, and R. S. Lee. 2008. Identifying unique devices through wireless fingerprinting. In *Proceedings of the First ACM Conference on Wireless Network Security - WiSec'08*, 46.
26. Cunche, M., Mohamed Ali Kaafar, and R. Boreli. 2012. I know who you will meet this evening! Linking wireless devices using wi-fi probe requests. In *2012 IEEE International Symposium on a World of Wireless, Mobile and Multimedia Networks (WoWMoM)*, 1–9.
27. Das, A., N. Borisov, and E. Chou. 2018. Every move you make: Exploring practical issues in smartphone motion sensor fingerprinting and countermeasures. *Proceedings on Privacy Enhancing Technologies* 2018 (1): 88–108.
28. Dey, S., N. Roy, W. Xu, R. R. Choudhury, and S. Nelakuditi. 2014. Accelprint: Imperfections of accelerometers make smartphones trackable. In *Network and Distributed System Security Symposium (NDSS)*.
29. Miluzzo, E., A. Varshavsky, S. Balakrishnan, and R. R. Choudhury. 2012. Tapprints: Your finger taps have fingerprints. In *MobiSys'12: Proceedings of the 10th International Conference on Mobile Systems, Applications, and Services, ACM* , 323.
30. Achara, J. P., G. Acs, and C. Castelluccia. 2015. On the unicity of smartphone applications. In *Proceedings of the 14th ACM Workshop on Privacy in the Electronic Society*, 27–36. New York: ACM.
31. Seneviratne, S., A. Seneviratne, P. Mohapatra, and A. Mahanti. 2014. Predicting user traits from a snapshot of apps installed on a smartphone. *ACM SIGMOBILE Mobile Computing and Communications Review* 18 (2): 1–8.
32. Bujlow, T., V. Carela-Español, J. Sole-Pareta, and P. Barlet-Ros. 2017. A survey on web tracking: Mechanisms, implications, and defenses. *Proceedings of the IEEE* 105 (8): 1476–1510.
33. Mayer, J. R., and J. C. Mitchell. 2012. Third-party web tracking: Policy and technology. In *2012 IEEE Symposium on Security and Privacy*, 413–427. Piscatawy: IEEE.

34. Ikram, M., H. J. Asghar, M. A. Kaafar, A. Mahanti, and B. Krishnamurthy. 2017. Towards seamless tracking-free web: Improved detection of trackers via one-class learning. *Proceedings on Privacy Enhancing Technologies* 2017 (1): 79–99.
35. Atterer, R., M. Wnuk, and A. Schmidt. 2006. Knowing the user's every move: User activity tracking for website usability evaluation and implicit interaction. In *Proceedings of the 15th International Conference on World Wide Web*, 203–212. New York: ACM.
36. Acar, G., M. Juarez, and N. Nikiforakis. 2013. FPDetective: Dusting the web for fingerprinters. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, 1129–1140.
37. Ullrich, J. B. 2015. 11 ways to track your moves when using a web browser. https://isc.sans.edu/forums/diary/11+Ways+To+Track+Your+Moves+When+Using+a+Web+Browser/19369/
38. Timber, C. 2014. Verizon, AT&T tracking their users with 'supercookies'. https://www.washingtonpost.com/business/technology/verizon-atandt-tracking-their-users-with-supercookies/2014/11/03/7bbbf382-6395-11e4-bb14-4cfea1e742d5_story.html
39. Sweeney, L. 2000. Simple demographics often identify people uniquely. Carnegie Mellon University, Data Privacy Working Paper 3. Pittsburgh 2000, 1–34.
40. Krishnamurthy, B., and C. Wills. 2009. Privacy diffusion on the web: a longitudinal perspective. In *Proceedings of the 18th International Conference on World Wide Web*, 541–550. New York: ACM.
41. Eckersley, P. 2010. How unique is your browser? In *Proceedings of the Privacy Enhancing Technologies Symposium (PETS)*, 1–18.
42. Yen, T.-F., Y. Xie, F. Yu, R. P. Yu, and M. Abadi. 2012. Host fingerprinting and tracking on the web: Privacy and security implications. In *Network and Distributed System Security Symposium*, 1–16.
43. Boda, K., A. M. Foeldes, G. G. Gulyas, and S. Imre. 2012. User tracking on the web via cross-browser fingerprinting. *Information Security Technology for Applications* 7161: 31–46.
44. Olejnik, Ł., G. Acar, C. Castelluccia, and C. Diaz. 2016. The leaking battery: A privacy analysis of the HTML5 battery status API. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 9481: 254–263.
45. Laperdrix, P., W. Rudametkin, and B. Baudry. 2016. Beauty and the beast: Diverting modern web browsers to build unique browser fingerprints. In *2016 IEEE Symposium on Security and Privacy (SP)*, 878–894. Piscataway: IEEE.
46. Fifield, D., and S. Egelman. 2015. Fingerprinting web users through font metrics. *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)* 8975: 107–124.
47. Englehardt, S., and A. Narayanan. 2016. Online tracking: A 1-million-site measurement and analysis. In *Proceedings of the 2016 ACM SIGSAC Conference on Computer and Communications Security - CCS'16*, 1388–1401.
48. Libert, T. 2015. Exposing the hidden web: An analysis of third-party HTTP requests on 1 million websites. *International Journal of Communication* 9, 3544–3561.
49. Mowery, K., and H. Shacham. 2012. Pixel perfect : Fingerprinting canvas in HTML5. In *Web 2.0 Security & Privacy 20 (W2SP)*, 1–12.
50. Weinberg, Z., E. Y. Chen, P. R. Jayaraman, and C. Jackson. 2011. I still know what you visited last summer: Leaking browsing history via user interaction and side channel attacks. In *2011 IEEE Symposium on Security and Privacy*, 147–161. Piscataway: IEEE.
51. Zalewski, M. 2008. Browser Security Handbook, Part 2. Mountain View: Google.
52. Chen, T., A. Chaabane, P. U. Tournoux, M.-A. Kaafar, and R. Boreli. 2013. How much is too much? Leveraging ADS audience estimation to evaluate public profile uniqueness. In *International Symposium on Privacy Enhancing Technologies Symposium*, 225–244. Berlin: Springer.

53. Perito, D., C. Castelluccia, M. A. Kaafar, and P. Manils. 2011. How unique and traceable are usernames? In *International Symposium on Privacy Enhancing Technologies Symposium*, 1–17. Berlin: Springer.
54. DeDeo, S. 2006. Pagestats. http://web.cs.wpi.edu/~cew/pagestats/.
55. Openwpm. 2019. https://github.com/mozilla/OpenWPM
56. Libert, T. webxray. https://webxray.org
57. Das, A., N. Borisov, and M. Caesar. 2016. Tracking mobile web users through motion sensors: Attacks and defenses. In *Network and Distributed System Security Symposium (NDSS)*.
58. Das, A., and N. Borisov. 2014. Poster: Fingerprinting smartphones through speaker. In *Poster at the IEEE Security and Privacy Symposium*. Princeton: Citeseer.
59. Das, A., N. Borisov, and M. Caesar. 2014. Do you hear what i hear? Fingerprinting smart devices through embedded acoustic components. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security*, 441–452. New York: ACM.
60. Zhou, Z., W. Diao, X. Liu, and K. Zhang. 2014. Acoustic fingerprinting revisited: Generate stable device ID stealthily with inaudible sound. In *Proceedings of the 2014 ACM SIGSAC Conference on Computer and Communications Security - CCS'14*, 429–440.
61. Bojinov, H., Y. Michalevsky, G. Nakibly, and D. Boneh. 2014. Mobile device identification via sensor fingerprinting. arXiv:1408.1416.
62. Corripio, J., D. González, A. Orozco, L. Villalba, J. Hernandez-Castro, and S. Gibson. 2013. Source smartphone identification using sensor pattern noise and wavelet transform. In *5th International Conference on Imaging for Crime Detection and Prevention, ICDP 2013*.
63. Lukáš, J., J. Fridrich, and M. Goljan. 2006. Digital camera identification from sensor pattern noise. *IEEE Transactions on Information Forensics and Security* 1 (2): 205–214.
64. Masood, R., B. Z. H. Zhao, H. J. Asghar, and M. A. Kaafar. 2018. Touch and you're trapp (CK) ed: Quantifying the uniqueness of touch gestures for tracking. *Proceedings on Privacy Enhancing Technologies* 2018 (2): 122–142.
65. Fredrikson, M., and B. Livshits. 2011. Repriv: Re-imagining content personalization and in-browser privacy. In *32nd IEEE Symposium on Security and Privacy, S&P 2011, 22–25 May 2011, Berkeley*, 131–146.
66. Toubiana, V., A. Narayanan, D. Boneh, H. Nissenbaum, and S. Barocas. 2010. Adnostic: Privacy preserving targeted advertising. In *Proceedings Network and Distributed System Symposium*.
67. Spooren, J., D. Preuveneers, and W. Joosen. 2015. Mobile device fingerprinting considered harmful for risk-based authentication. In *Proceedings of the Eighth European Workshop on System Security*, 6. New York: ACM.
68. Simon, L., W. Xu, and R. Anderson. 2016. Don't interrupt me while i type: Inferring text entered through gesture typing on android keyboards. *Proceedings on Privacy Enhancing Technologies* 2016 (3): 136–154.
69. Kurtz, A., H. Gascon, T. Becker, K. Rieck, and F. Freiling. 2016. Fingerprinting mobile devices using personalized configurations. *Proceedings on Privacy Enhancing Technologies* 2016 (1): 4–19.
70. Zhao, S., J. Ramos, J. Tao, Z. Jiang, S. Li, Z. Wu, G. Pan, and A. K. Dey. 2016. Discovering different kinds of smartphone users through their application usage behaviors. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing - UbiComp'16*, 498–509.
71. Dey, S. 2014. Accelprint: Data and source code. http://sdey4.web.engr.illinois.edu/AccelPrintDataSourceCode.html
72. Das, A. 2018a. Collecting sensor data from smart devices. https://anupamdas.org/SensorDataCollection.html
73. Das, A. 2018b. Fingerprinting smartphones via microphones and speakers. https://anupamdas.org/acoustistic_fp.html
74. Kohno, T., A. Broido, and K. C. Claffy. 2005. Remote physical device fingerprinting. *IEEE Transactions on Dependable and Secure Computing* 2 (2): 93–108.

75. Pang, J., B. Greenstein, R. Gummadi, S. Srinivasan, and D. Wetherall. 2007. 802. 11 user fingerprinting. *Proceedings of the 13th Annual ACM International Conference on Mobile Computing and Networking* 9: 99–110.
76. Hall, J., M. Barbeau, and E. Kranakis. 2003. Detection of transient in radio frequency fingerprinting using signal phase. *Wireless and Optical Communications*, 13–18.
77. Nguyen, N. T., G. Zheng, Z. Han, and R. Zheng. 2011. Device fingerprinting to enhance wireless security using nonparametric Bayesian method. In *Proceedings of IEEE Infocom*, 1404–1412.
78. Acar, G., M. Juarez, N. Nikiforakis, C. Diaz, S. Gürses, F. Piessens, and B. Preneel. 2013. Fpdetective: Dusting the web for fingerprinters. In *Proceedings of the 2013 ACM SIGSAC Conference on Computer & Communications Security*, 1129–1140. New York: ACM.
79. Ruotsalo, T., K. Athukorala, D. Głowacka, K. Konyushkova, A. Oulasvirta, S. Kaipiainen, S. Kaski, and G. Jacucci. 2013. Supporting exploratory search tasks with interactive user modeling. *Proceedings of the American Society for Information Science and Technology* 50 (1), 1–10.
80. Ha, I., K.-J. Oh, and G.-S. Jo. 2015. Personalized advertisement system using social relationship based user modeling. *Multimedia Tools and Applications* 74 (20): 8801–8819.
81. Xue, B., L. Wu, K. Wang, X. Zhang, J. Cheng, X. Chen, and X. Chen. 2021. Multiuser gesture recognition using semg signals via canonical correlation analysis and optimal transport. *Computers in Biology and Medicine* 130: 104188.
82. Magrabi, F., I. Habli, M. Sujan, D. Wong, H. Thimbleby, M. Baker, and E. Coiera. 2019. Why is it so difficult to govern mobile apps in healthcare? *BMJ Health and Care Informatics* 26 (1): e100006.
83. Chellappa, R. K., and R. G. Sin. 2005. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management* 6 (2–3): 181–202.
84. Berkovsky, S., N. Borisov, Y. Eytani, T. Kuflik, and F. Ricci. 2007. Examining users' attitude towards privacy preserving collaborative filtering. In *Workshop on Data Mining for User Modeling, Online Proceedings*, 28.
85. Masood, R., D. Vatsalan, M. Ikram, and M. A. Kaafar. 2018. Incognito: A method for obfuscating web data. In *WWW '18: Proceedings of the 2018 World Wide Web Conference*, 267–276.
86. Toch, E., Y. Wang, and L. F. Cranor. 2012. Personalization and privacy: A survey of privacy risks and remedies in personalization-based systems. *User Modeling and User-Adapted Interaction* 22 (1–2): 203–220.
87. Morrissey, P. 2018. 6 people who were fired for social media posts. https://www.smithslawyers.com.au/post/6-people-who-were-fired-for-social-media-posts
88. Schroeder, S. 2009. Domino's youtube video: Youtube can get you fired, too. https://mashable.com/2009/04/14/youtube-fired/
89. Ostrow, A. 2009. Facebook fired: 8% of us companies have sacked social media miscreants. https://mashable.com/2009/08/10/social-media-misuse/.
90. Ullah, I., R. Boreli, M. A. Kaafar, and S. S. Kanhere. 2014. Characterising user targeting for in-app mobile ads. In *2014 IEEE Conference on Computer Communications Workshops (INFOCOM WKSHPS)*, 547–552. Piscataway: IEEE.
91. United states - google transparency report. 2014. https://transparencyreport.google.com/user-data/us-national-security?hl=en
92. Post, T. W. 2013b. NSA uses google cookies to pinpoint targets for hacking. http://www.washingtonpost.com/blogs/the-switch/wp/2013/12/10/nsa-uses-google-cookies-topinpoint-targets-for-hacking
93. Post, T. W. 2013a. NSA tracking cellphone locations worldwide, snowden documents show. https://www.washingtonpost.com/world/national-security/nsa-tracking-cellphone-locations-worldwide-snowden-documents-show/2013/12/04/5492873a-5cf2-11e3-bc56-c6ca94801fac_story.html

94. Duhigg, C. 2012. How companies learn your secrets. https://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all
95. Singel, R. 2008. Facebook beacon tracking program draws privacy lawsuit. https://www.wired.com/2008/08/facebook-beacon/
96. Hansell, S. 2006. Aol removes search data on vast group of web users. http://query.nytimes.com/gst/fullpage.html?res=9504e5d81e3ff93ba3575bc0a9609c8b63
97. Narayanan, A., and V. Shmatikov. 2008. Robust de-anonymization of large sparse datasets. In *Proceedings of the 2008 IEEE Symposium on Security and Privacy, SP'08, Washington*, 111–125. Piscataway: IEEE.
98. Sweeney, L. 1997. Weaving technology and policy together to maintain confidentiality. *The Journal of Law, Medicine & Ethics* 25 (2–3): 98–110.
99. Canny, J. 2002. Collaborative filtering with privacy via factor analysis. In *Proceedings of the 25th Annual International ACM SIGIR Conference on Research and Development in Information Retrieval*, 238–245. New York: ACM.
100. Schafer, J. B., D. Frankowski, J. Herlocker, and S. Sen. 2007. Collaborative filtering recommender systems. In *The Adaptive Web*, 291–324. Berlin: Springer.
101. Canny, J. F. 2002. Collaborative filtering with privacy. In *2002 IEEE Symposium on Security and Privacy, Berkeley, California, May 12–15, 2002*, 45–57.
102. Nissim, K., T. Steinke, A. Wood, M. Altman, A. Bembenek, M. Bun, M. Gaboardi, D. R. O'Brien, and S. Vadhan. 2017. Differential privacy: A primer for a non-technical audience. In *Privacy Law Scholars Conference*.
103. Dwork, C., A. Roth, et al. 2014. The algorithmic foundations of differential privacy. *Foundations and Trends® in Theoretical Computer Science* 9 (3–4): 211–407.
104. Samarati, P., and L. Sweeney. 1998. Protecting privacy when disclosing information: k-anonymity and its enforcement through generalization and suppression. Technical Report, SRI International.
105. Chen, T., R. Boreli, M. A. Kâafar, and A. Friedman. 2014. On the effectiveness of obfuscation techniques in online social networks. In *Privacy Enhancing Technologies - 14th International Symposium, PETS 2014, Amsterdam, July 16–18, 2014. Proceedings*, 42–62.
106. Salamatian, S., A. Zhang, F. du Pin Calmon, S. Bhamidipati, N. Fawaz, B. Kveton, P. Oliveira, and N. Taft. 2013. How to hide the elephant- or the donkey- in the room: Practical privacy against statistical inference for large data. In *IEEE Global Conference on Signal and Information Processing, GlobalSIP 2013, Austin, December 3–5, 2013*, 269–272.
107. Li, C., H. Shirani-Mehr, and X. Yang. 2007. Protecting individual information against inference attacks in data publishing. In *Proceedings of the 12th International Conference on Database Systems for Advanced Applications, DASFAA'07, Berlin*, 422–433. Berlin: Springer.
108. Raval, N., A. Machanavajjhala, and J. Pan. 2019. Olympus: Sensor privacy through utility aware obfuscation. *Proceedings on Privacy Enhancing Technologies* 2019 (1): 5–25.
109. Bindschaedler, V., and R. Shokri. 2016. Synthesizing plausible privacy-preserving location traces. In *2016 IEEE Symposium on Security and Privacy (SP)*, 546–563. Piscataway: IEEE.
110. Cerf, S., V. Primault, A. Boutet, S. B. Mokhtar, R. Birke, S. Bouchenak, L. Y. Chen, N. Marchand, and B. Robu. 2017. Pulp: Achieving privacy and utility trade-off in user mobility data. In *2017 IEEE 36th Symposium on Reliable Distributed Systems (SRDS)*, 164–173. Piscataway: IEEE.
111. Boutet, A., and M. Cunche. 2018. A privacy-preserving mechanism for requesting location data provider with wi-fi access points. *International Journal of Applied Engineering Research*,12 (9): 1982–1986
112. Das, A., G. Acar, N. Borisov, and A. Pradeep. 2018. The web's sixth sense: A study of scripts accessing smartphone sensors. In *Proceedings of the 2018 ACM SIGSAC Conference on Computer and Communications Security*, 1515–1532. New York: ACM.
113. Arlein, R. M., B. Jai, M. Jakobsson, F. Monrose, and M. K. Reiter. 2000. Privacy-preserving global customization. In *Proceedings of the 2nd ACM Conference on Electronic Commerce*, 176–184. New York: ACM.

114. Hitchens, M., J. Kay, B. Kummerfeld, and A. Brar. 2005. Secure identity management for pseudo-anonymous service access. In *International Conference on Security in Pervasive Computing*, 48–55. Berlin: Springer.
115. Gerber, S., M. Fry, J. Kay, B. Kummerfeld, G. Pink, and R. Wasinger. 2010. Personisj: Mobile, client-side user modelling. In *International Conference on User Modeling, Adaptation, and Personalization*, 111–122. Berlin: Springer.
116. Kay, J. 2006. Scrutable adaptation: Because we can and must. In *International Conference on Adaptive Hypermedia and Adaptive Web-Based Systems*, 11–19. Berlin: Springer.
117. Kay, J., B. Kummerfeld, and P. Lauder. 2003. Managing private user models and shared personas. In *UM03 Workshop on User Modeling for Ubiquitous Computing*, 1–11. Princeton: Citeseer.
118. Online privacy tools for the general public. 2015. White paper, European Union Agency For Network and Information Security (ENISA).
119. Electronic Frontier Foundation (EFF). Do not track (DNT). https://www.eff.org/issues/do-not-track
120. Howe, D. C., and H. Nissenbaum. 2009. Trackmenot: Resisting surveillance in web search. *Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society* 23: 417–436.
121. Domingo-Ferrer, J., A. Solanas, and J. Castellà-Roca. 2009. h (k)-private information retrieval from privacy-uncooperative queryable databases. *Online Information Review* 33 (4): 720–744.
122. Shapira, B., Y. Elovici, A. Meshiach, and T. Kuflik. 2005. PRAW - A privacy model for the web. *Journal of the American Society for Information Science and Technology (JASIST)* 56 (2): 159–172.
123. Brave. 2019. https://brave.com/