Privacy considerations of online behavioural tracking

Contributors to this report

Authors:

- Claude Castelluccia (Inria, France)
- Arvind Narayanan (Stanford University, USA)

ENISA project management:

- Rodica Tirtea
- Demosthenes Ikonomou

Acknowledgements

The authors would like to thank their colleagues at Inria and Stanford for their valuable feedback and comments on this document. They are particularly grateful to Jonathan Mayer, Stanford, for his contributions in Section 5 and Section 8.2. Some text in Sections 5 and 8.2 was excerpted from Mayer and Mitchell [Maye2012] with permission.

[Deliverable – 2012-10-19]

About ENISA

The European Network and Information Security Agency (ENISA) is a centre of network and information security expertise for the EU, its Member States, the private sector and Europe's citizens. ENISA works with these groups to develop advice and recommendations on good practice in information security. It assists EU Member States in implementing relevant EU legislation and works to improve the resilience of Europe's critical information infrastructure and networks. ENISA seeks to enhance existing expertise in EU Member States by supporting the development of cross-border communities committed to improving network and information security throughout the EU. More information about ENISA and its work can be found at www.enisa.europa.eu.

Contact details

For contacting ENISA or for general enquiries on this subject, please use the following details:

- Email: sta@enisa.europa.eu
- Internet: http://www.enisa.europa.eu

For questions related to this project, please use the following details:

- Email: sta@enisa.europa.eu

Follow ENISA on Facebook, Twitter, LinkedIn YouTube & RSS feeds

Contents

1 Executive Summary

Internet users are being increasingly tracked[1] and profiled and their personal data are extensively used as currency in exchange for services. It is important that this new reality is better understood by all stakeholders if we are to be able to support and respect the right for privacy. This study complements previous and current work of the European Network and Information Security Agency (ENISA) in the area of privacy and data protection. ENISA is supporting a dialogue between various stakeholders to identify challenges, develop best practice and promote privacy concepts and technologies in the operational environment.

In recent studies[2] published by ENISA we observed that there is a gap between the legal requirements for personal data protection, i.e. minimal disclosure principle and minimum duration of the storage of personal data, and the practice in the online environment. The uptake of privacy-enhancing technologies is low. Users do not have many options, even if privacy-friendly services could bring some business advantages, as a small but not insignificant proportion of consumers are even willing to pay for privacy-friendly services.[3]

The new Regulation[4] proposal of the European Commission, where privacy-by-design and by default is promoted and sanctions are included, aims to address these challenges from a legal perspective.

This study provides a technical perspective on behavioural tracking. It presents a comprehensive view, answering questions such as: Why are users tracked? What techniques are used? To what extent are we tracked today? What are the trends? What are the risks? What protective measures exist? What could regulators do to help improve user privacy?

More work is needed and in an interdisciplinary approach to address the privacy risks associated with tracking mechanisms. The recommendations of this study are addressed to regulators, policy stakeholders, researchers and developers. They are as follows:

---

[1] *'Cookie counts continue to increase, with larger and larger amounts of third party cookies being used. Cookies are present on every website in the top 100 […]. Third party trackers continue to increase and we expect this trend to continue as well.' Nathan Good & Chris Jay Hoofnagle, The Web Privacy Census, June 2012, available at http://law.berkeley.edu/privacycensus.htm*

[2] *ENISA, Study on data collection and storage in the EU, February 2012, available at: http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/data-collection*

ENISA is in the process of publishing a new study by the end of 2012: 'The right to be forgotten – between expectations and practice'. It will be available under http://www.enisa.europa.eu/act/it/library/

[3] *ENISA, Study on monetising privacy. An economic model for pricing personal information, February 2012, available at: http://www.enisa.europa.eu/activities/identity-and-trust/library/deliverables/monetising-privacy*

[4] *European Commission, Proposal for a regulation of the European Parliament and of the Council on the protection of individuals with regard to the processing of personal data and on the free movement of such data (General Data Protection Regulation), COM(2012) 11 final, 25 January 2012, available at http://ec.europa.eu/justice/data-protection/document/review2012/com_2012_11_en.pdf (last accessed on 04.07.2012)*

- Development of anti-tracking initiatives and solutions for mobile applications; the users of mobile devices are more exposed as most anti-tracking initiatives are not focusing on mobile devices
- Development of easy-to-use tools for transparency and control; awareness is important but there is a need to enhance transparency tools to allow the users to know how their personal data is collected, managed and transferred
- Enforcement solutions should be deployed to block misbehaving players and to force compliance with rules and regulations regarding personal data protection; mechanisms should be defined by regulatory bodies both for compliance and for monitoring and detection of violation of the rules
- Privacy-by-design should be promoted; regulations have an important role in boosting the adaptation of privacy-preserving solutions, i.e. by enforcing the rules, and by ensuring the existence of complete, compliant, concrete and meaningful privacy policies.

2 Introduction

While the debate over online behavioural advertising and tracking has been going on for several years, it has recently intensified due to media coverage – for example, the *Wall Street Journal* 'What They Know' series [WSJ]. The main goal of this report is to present a state-of-the-art of *behavioural tracking and profiling* on the Internet, and to highlight some of the resulting potential privacy threats.

Behavioural targeting is the practice of tailoring online content, especially advertisements, to visitors based on their inferred interests, or 'profile'. The process of constructing this profile using data mining – transforming data into knowledge – is known as *online behavioural profiling*. The underlying data is typically a log of the user's web activity, and the data collection process is called *behavioural tracking*.

Tracking is categorised into *first-party* and *third-party* tracking. In *first-party tracking*, the tracking is performed by the site or application with which the user is directly interacting. In *third-party tracking*, the tracking is performed by other 'third party' entities, different from the entity the user is directly connected to (the user being the 'second party'), that track the user's browsing activity over time and across different websites. For example, Facebook tracks across sites via its 'Like' button; each time a user visits a site that contains a Facebook 'Like' button, Facebook is informed about it, even if the user does not click on this button.

Internet users are being increasingly tracked and profiled [Krish2009b, Krish2009c, Chab2012a, Chab2012b]. Companies use profiling to provide customised, i.e. personalised, services to their customers, with the goal of increasing revenues. In particular, behavioural advertising takes advantage of profiles of users' interests, characteristics, such as age and gender, and purchasing activities. Advertising or publishing companies use behavioural targeting to display advertisements that closely reflect users' interests, e.g. 'sports enthusiasts'. It can be argued that customisation resulting from profiling is also beneficial to users, who receive useful information and relevant online ads in line with their interests.

However, behavioural tracking is often perceived as a threat to privacy, mainly because it relies heavily on users' personal information. One possible negative consequence is a surveillance society or Internet, where all our online or physical activities are recorded and correlated.

This report is structured as follows: Section 3 presents the motivations behind tracking. Section 4 lists the main tracking techniques. Section 5 discusses how tracking is performed on the Internet today. Section 6 looks into the future of online tracking. Section 7 discusses the dangers of tracking. Section 8 presents different countermeasures that are currently being proposed to mitigate the effects of tracking. Finally, section 9 concludes this report with some recommendations.

3 Motivations: Why are Internet users tracked?

There are a variety of motivations behind online tracking. First-party tracking is often performed by website owners to personalise user experience across sessions, such as maintaining the user's shopping cart and preferences. First-party tracking is also used for fraud detection and law enforcement. In fact, several regulations require websites to log users' activities for the purpose of fraud prevention, anti-money laundering, national security and law enforcement [Tene2011]. Two major reasons for third-party tracking are user profiling, which is used in targeted advertising, and measurement/analytics. These two aspects are detailed in the rest of this section.

User profiling. The intention of behavioural targeting is to track users over time and build profiles of their interests, characteristics, such as age and gender, and shopping activities. Online advertisements use behavioural targeting to display advertisements that reflect users' interests. To a first approximation, online advertising systems are composed of three main entities: the *advertiser*, the *publisher* and the *ad network*. The advertiser is the entity, such as a car manufacturer or a hotel, which wants to advertise a product or service. The publisher is the entity, such as an online newspaper company, that owns one or several websites and is willing to display advertisements and be paid for it. Finally, the ad network is the entity that collects advertisements from the advertisers and places them on publisher sites. If a user clicks on an advertisement (in the 'cost-per-click' model), the ad network collects payment from the corresponding advertiser, and pays out a part of it to the publisher. There is, therefore, a strong incentive for the ad network to generate accurate and complete profiles in order to maximise the 'click-through rate' and consequently revenues.

E-commerce sites, in the first-party context, also use behavioural tracking and profiling to recommend products that are likely to be of interest to users. For example, Amazon recommends products to online users based on individuals' past behaviours (*personalised recommendation*), on past behaviours of similar users (*social recommendation*) and, of course, on searched items (*item recommendation*) [Macm2009].

With the emergence of smartphones, many applications record users' locations and movement. Location information enables many useful services such as driving directions, knowing where their friends are or recommendations for nearby restaurants. However, this information is also collected by marketers to improve profiling. While the benefits provided by these systems are indisputable, they unfortunately pose a considerable threat to location privacy, as illustrated by the recent iPhone and Android controversies [Raph2011].

Web analytics/measurement. Tracking is also used for various types of aggregate measurements, such as website traffic statistics or effective exposure of advertising [Tene2011]. Although it is technically feasible for first parties to carry out these measurements on their own, many websites use third-party web analytics tools, such as *Google analytics* [GoAn], to obtain aggregate traffic statistics such as most visited pages, visitors' countries, etc. These tools typically track users to collect their browsing activities and

to periodically compile them into aggregated statistics. These statistics are often used by websites to measure the effectiveness of ad campaigns or to optimise their content.

4 A quick survey of existing tracking techniques

Online tracking technologies have evolved considerably in the last few years [Ecke2009, Scho2009]. In this section, we present the main online tracking technologies. A more detailed description of how tracking is actually implemented on the web can be found in [Roes2012]. A survey of both technological and policy aspects of tracking was recently published at the IEEE Security&Privacy conference [Maye2012].

One of the main sources of information used for profiling comes from web tracking, i.e., tracking users across different visits or across different sites. Data collected include the sequence of visited sites and viewed pages, and the time spent on each page. Web tracking is mainly performed by monitoring IP addresses and cookies, using techniques such as Javascript, supercookies, fingerprinting, or DPI (Deep Packet Inspection). The latter is used by some ISPs and this practice remains controversial.[5] With the emergence of smartphones, equipped with more and more sophisticated sensors, location and physical activities are also becoming important sources of information for profiling.

Cookies. A cookie is a piece of text stored by a user's web browser and transmitted as part of an HTTP request. It consists of one or more name-value pairs containing bits of information and is set by a web server. There are two types of cookies: *session* and *persistent* cookies. Session cookies are temporary cookies that are often used to store user choices or navigation state. They are set by a service when a user logs in, and are erased when the user logs out. Persistent cookies are often used to store identifying information, user preferences or authentication tokens to keep an authenticated session with a server. These files stay in the user's browser until they are explicitly deleted or they expire. They are sent back unchanged by the browser each time it accesses that website and can, therefore, be used by websites to track users across visits. Persistent cookies raise serious privacy concerns. They are sent only to the websites that set them or to servers in the same domain. However, a web page may contain images, links, web bugs,[6] HTML IFrame, JavaScript, or other components stored on servers in other domains. Cookies that are set during retrieval of these components are called third-party cookies, in contrast to first-party cookies. Some sites, such as advertising companies, use third-party cookies to track users across multiple sites. In particular, an advertising company can track a user across all pages where it has placed advertising images or web bugs. Knowledge of the pages visited by a user allows the advertising company to target advertisements to users' presumed preferences.[7]

Javascript. Many websites contain executable Javascript files that are downloaded by visiting users. These files sometimes update first-party cookies and send information back to the

[5] *DPI is also addressed by Directive 2009/136/EC (also known as the ePrivacy directive).*

[6] *Objects (i.e. 1x1 pixel GIF images) present on a web page or in an email, usually invisible to the user, that allow checking whether a user has viewed the page or email.*

[7] *For a concise presentation on how cookies work, the most common types of cookies and an analysis of the respective security vulnerabilities and privacy concerns, see the paper published by ENISA entitled 'Bittersweet cookies. Some security and privacy considerations' (Feb 2011), available at: http://www.enisa.europa.eu/activities/identity-and-trust/library/pp/cookies/*

servers. Javascript programs have limited access to user data. However, they can access information stored in the browser including cached objects and the history of visited links.

Along with cookies and results of JavaScript execution, trackers have all the regular information available in a typical HTTP request, unless the user has explicitly taken steps to block some of it: the user's IP address, the user-agent string (i.e., information about the browser and possibly add-ons), current and previous URL (via Referer header), language preference (Accept-Language header), etc.

Supercookies and Evercookies. Use of tracking cookies is fairly ubiquitous and there are known techniques to avoid them [Dixo2011]. Therefore, there is an impetus in the Internet tracking industry to discover and deploy more robust tracking mechanisms, often referred to as Supercookies [McKi2008]. One of the most prominent supercookies is the so-called 'Flash cookie', a type of cookie maintained by the Adobe Flash plugin on behalf of Flash applications embedded in web pages [Scho2009]. Since these cookie files are stored outside of the browser's control, web browsers did not traditionally provide an interface to view, manage and delete these cookies. In particular, users are not notified when such cookies are set, and these cookies never expire. Flash cookies can track users in all the ways traditional HTTP cookies do, and they can be stored or retrieved whenever a user accesses a page containing a Flash application. Flash cookies are extensively used by popular sites. They are often used to circumvent users' HTTP cookie policies and privacy preferences. For example, it was found that some sites use HTTP and Flash cookies that contain redundant information [Ashk2009]. Since flash cookies do not expire, sites might automatically re-spawn HTTP cookies from Flash ones if they are deleted. The persistence of supercookies can be further improved [Kamk2010]. This new type of cookie, called *evercookie*, is a combination of various tracking mechanisms, each reinforcing the others, and is able to identify a client even when standard cookies and Flash cookies have been removed.

Stateless tracking (Browser fingerprinting). A recent study showed that browsers can be identified to a high degree of accuracy without cookies or other tracking technologies [Ecke2010]. Web browsers provide various pieces of information to websites, such as User Agent, fonts, screen resolution, etc., that may not be capable of identifying a browser on their own but are capable of doing so when used in combination. The study shows that a browser fingerprint is unique enough so that it can, on the average, identify a browser among a set of 290,000 other browsers (this is a conservative estimate of uniqueness). Browser fingerprinting is a powerful tool for tracking users along with IP addresses, cookies and supercookies. This type of tracking, called stateless or passive tracking, is particularly problematic since it is hard to detect.

Location tracking. The W3C geolocation API, which is supported in the Firefox, Opera and Chrome browsers and in Internet Explorer via a plugin, allows websites to request geographical information for the client device. With the approval of the user, the browser sends information like the client's IP address, MAC addresses of connected wireless access points and the cell ids of GSM/CDMA networks within range. With the help of a network location provider, such as Google Location Services, this information can be used to obtain an

estimate of the client devices location. While the browser only sends this information to a website with user explicit approval, few users realise the accuracy with which these services can often locate a device. For instance, Google Location Services rely on the MAC addresses of wireless access points detected during the Google Street View data collection to locate client devices to within the range of an IEEE 802.11 wireless base station (i.e., tens of metres).

5 What do they know about us?

The previous section listed some of the existing tracking techniques. In this section, we discuss how these techniques are used by marketers, social networks, and smartphone applications, to track and profile users.

Third-party tracking. The increasing presence and tracking of third-party sites used for advertising and analytics has been demonstrated in a study [Krish2009b, Krish2009c]. This study showed that the penetration of the top 10 third parties grew from 40% in 2005 to 70% in 2008, and to over 70% in September 2009. Another study shows that not only are these third parties increasing their tracking of users, but that they can now link these traces with identifiers and personal information via online social networks [Krish2009a]. However, it was recently shown that the combination of UA and IP prefix (not even full address) can be used to identify a host with a probability of 95%. This suggests that anonymisation techniques that store the IP prefix do not provide much privacy. Cookie IDs offer only slightly better performance than the use of US and IP prefixes [Yen2012].

Online Social Network (OSN) tracking. Most popular social networking websites, such as Facebook, Twitter, Xing and Google+, track users around the web. Each of these social networks have social widgets for sharing and recommendation (called *Like*, *Tweet*, *Visitors*, and *+1* buttons) which are installed on numerous websites. These buttons allow the social networks to track users, even when they don't click those buttons – just viewing a webpage with such a button is sufficient to be tracked.

A recent study showed that out of the 10K most popular websites (according to Alexa ranking), 22% contain a Facebook 'Like' button, 7.5% a 'Twitter Re-Tweet' button, and 10.4% contain a 'Google+ share' button. More seriously from a privacy perspective, this study showed that 22 out of the 77 health-related sites that appear among these 10K sites contain a Facebook 'Like' button. Unsurprisingly, these percentages only increase with time [Chab2012a, Chab2012b]. Note that these social networks are able to track users who are not logged in. Furthermore, this tracking is possible even if the user does not participate in the social network, i.e., does not have an account, as long as he or she has visited the social network at least once (i.e. has a cookie set by the social network). In the latter scenario, the social network does not learn the identity of the user. However, the tracking logs could potentially be associated with an identity if and when the user creates an account using the same browser.

Mobile device tracking. Hundreds of millions of people worldwide use at least one smartphone. These mobile phones have increasing computational capacities and are equipped with multiple sensors like microphones, cameras, GPS, accelerometers, etc. They also contain a lot of personal information about their owners: phone numbers, current location, the owner's real name, a unique phone ID number. More and more geolocated applications enable individuals and communities to collect and share various kinds of data.

Most users remain unaware of the extra information that is collected about them beyond explicitly requested data. A recent *Wall Street Journal* study [Thur2010] showed that several

of the most popular Android or iPhone applications, including games and OSNs, transmitted the phone's unique device ID, phone's location, age, gender and other personal details to third-party companies without users' awareness or consent. The privacy risk becomes higher as the boundary between OSN and Location-Based Services (LBS) becomes fuzzier. For instance, OSNs such as FourSquare and Facebook are designed to encourage their users to share their geolocated data, and information posted on social applications such as Twitter can be used to infer whether or not an individual is at home. Other applications, such as Google Latitude, allow users to track the movements of their friends' cellphones and display their position on a map. In addition to social applications, there are other public sources of information that can be utilised by potential adversaries, such as the free geographic knowledge provided by Google Maps, Yahoo! Maps and Google Earth.

Re-identification: It is often argued that most of the tracking described above is harmless, because traces are anonymous. In other words, although sites are able to track devices, they cannot tell who the users behind them are. Of course, things are not that simple in practice. A trace can often be deanonymised and linked to an identity via different methods. Narayanan [Nara2011a] recently proposed a taxonomy of several ways in which a pseudonymous browsing history might become identified.

1. The third party is also a first party: The third party may be a first party in another context, where the user voluntarily provided her identity. Facebook, for example, has over 800 million users and enforces a requirement that users provide their real name to the service. When a page includes a third-party Facebook social widget, Facebook identifies the user to personalise the widget.

2. A first party sells the user's identity: Some first-party websites intentionally provide a user's identity to third parties if they are paid. Some have even made a business model of it, usually appearing as a free sweepstakes or quiz. Several advertising data providers buy identifying information, retrieve the user's dossier from an offline consumer database, and use it to target advertising.

3. A first party unintentionally provides ('leaks') identity: If a website puts identifying information in a URL or page title, it may unintentionally leak the information to third parties. In a 2011 paper [Krish2011], Krishnamurthy et al. examined signup and interaction with 120 popular sites for information leakage to third parties. They reported that an aggregate of 48% leaked a user identifier in a Request-URI or referrer.

4. De-anonymisation: The third party could match pseudonymous browsing histories against identified datasets to re-identify them. Re-identification of longitudinal, 'high-dimensional' data has been demonstrated in various contexts such as by Narayanan and Shmatikov on the Netflix Prize dataset [Nara2008].

Furthermore, users participate in different sites and leave piece of information (online social footprints) about themselves on many of them. This information is often public and can easily be collected to build profiles. One challenge here is to put all the pieces together, i.e., to link the different public but pseudonymous online profiles of a single

user, given that users typically register with different pseudonyms on different services. However, it was recently shown that a significant portion of the users choose a small number of related and predictable usernames and use them across many services [Peri2011].

There is tremendous commercial value in linking together every piece of online information about an individual. While the academic study linkage of social profiles is new, commercial firms have long been scraping profiles, aggregating them, and selling them on the grey market. Well-known public-facing aggregators such as Spokeo mainly use public records, but online profiles are quickly becoming part of the game [Nara2011c].

6 Future trends

Tracking techniques have evolved significantly over the past few years, and will continue to evolve. In the section, we discuss some of the possible future tracking trends.

Reality/Physical mining. Reality mining infers human relationship and behaviour from information collected by smartphones [Gree2008]. This information includes data collected by cellphone sensors, such as location or physical activity, and data recorded by phones themselves, such as the duration of the calls and the numbers dialled. Reality mining could help users identify things to do or new people to meet. It could also help to monitor health. For example, monitoring a phone's motion might reveal changes in gait, which could be an early indicator of ailments or depression. The idea of autonomous search is a first step toward reality mining. With autonomous search, the search engine will conduct searches for users without them having to manually type anything [Boult2010]. For example, a user could be walking down a street and receive personalised information about the places in the vicinity on his or her mobile phone, without having to click any buttons. While the promise of reality mining is great, the idea of collecting so much personal information naturally raises many questions about privacy and portends the spectre of a surveillance society.

Augmented reality. In a recent and fascinating study, Acquisti and his CMU colleagues showed that the convergence of face recognition, social networks, data mining, and cloud computing can be used to link offline and online public data to recover very sensitive information about a person [Acqu2011]. They first showed that face-recognition tools could be used to re-identify anonymous online profiles. They took unidentified profile photos from a popular dating site, where people use pseudonyms to protect privacy. They then compared these photos, using face recognition, to photos available on Facebook public profiles, and showed that they were able to re-identify a significant proportion of members of the dating site. Second, they show that it is possible to obtain the identity of strangers in the street. They took photos of strangers with a webcam and compared them to images from Facebook profiles. Using this approach, they re-identified about one-third of the subjects in the experiment. Finally, they show that it is possible to predict the interests and few digits of the Social Security numbers of some of the participants in the second experiment from their face and information retrieved from their OSN sites. This study highlights serious privacy concerns raised by the convergence of various technologies. With the improvement of data mining technology, such inference techniques will become increasingly feasible. Furthermore, it is not clear how self-regulation, opt-in mechanisms, or even regulation can help prevent this type of disclosure since all presented results were based on publicly available information.

7 The risks of tracking: What are the dangers of tracking?

In this section we discuss some concrete examples of the risks and dangers of online tracking.

7.1 Surveillance (government, companies)

One of the biggest risks of tracking is global surveillance. This surveillance can be performed by government, for security or political reasons, or by companies for commercial reasons. As detailed in a *New York Times* article [NYT2012], marketers have long understood the benefits of learning and influencing consumers' habits. Detecting major changes in behaviour increases the odds of getting customers to switch to a different product. This monitoring was previously performed via different types of fidelity cards. Internet tracking is a more powerful tool since it allows marketers to adapt their strategies almost instantaneously. As shown in [NYT2012], marketers use prediction models that can tell from a user's change of behaviour if she is pregnant or getting divorced. Although tracking has huge economic benefits, it raises serious privacy concerns.

Companies often advance the 'nothing-to-hide' argument to justify their activities – why would a user be concerned about his privacy if he has nothing to hide? Solove refutes this argument by pointing out that it stems from a narrow conception of privacy as secrecy or concealment of information [Solo2011]. Solove also notes that privacy dangers do not necessarily manifest as visceral injuries or damage. Information-gathering programs are problematic even if no information that people want to hide is uncovered. Collected information can be incorrect or distorted, and result in incorrect decisions, which will create frustration. The potential harms are error, abuse, lack of transparency and accountability.

7.2 Service discrimination and price discrimination

Another consequence of tracking and profiling is service discrimination or exclusion. Profiling may reveal that a user is suffering from, or has a propensity to develop, a certain disease. This information could, for example, be used by a health insurance firm to deny insurance or to significantly increase premiums. Price discrimination has a long history and is a common practice today [Nara2011b]. However, currently it is typically carried out via an explicit attribute of a buyer such as his age or gender. With tracking and profiling, service and price discrimination may be customised to each individual. Traditionally, there has never been enough data to do this.

7.3 The risks of personalisation

As described previously, profiling is often used by service providers to personalise their content to users. A news site may display only news matching users' previous reading patterns. A merchant site may propose only products that match the user's inferred interests, needs or preferences. Search engines may refine results based on a user's previous queries and clicks. And of course, online advertisements are often behaviourally targeted. This personalisation is a cause for concern. As argued by Eli Pariser, with service personalisation,

users get trapped in a 'filter bubble' and don't get exposed to information that could broaden their worldview [Pari2011]. In authoritarian states, personalisation could also be used to increase censorship by selecting news to show to specific users.

Conversely, content and service personalisation can be a source of information leakage, as it is often possible to retrieve a user's interests from the content/services provided to him using various inference techniques. For example, it was shown that a user's google history can be partially reconstructed from his query recommendations and that a user's interest profile can be inferred from his targeted ads [Cast2010, Cast2012]. In another example, a man discovered his teenage daughter was pregnant because he received coupons for baby food from the US superstore Target. The teenager had been profiled as pregnant from her purchase behaviour [NYT2012].

8 Protective measures. What can be done to mitigate tracking/profiling?

As illustrated in this report, users are being constantly tracked and profiled when using the Internet. This section discusses existing technological, legislative and educational protective measures.

8.1 Technological measures

Visualisation and blocking tools: There are several browser plugins, such as *Collusion* or *PrivacyBucket*, that show users how much trackers may be able to learn about them. There are also many browser tools and plugins that detect and block all or a list of third-party trackers. For example, NoScript is a Firefox add-on that allows executable content such as JavaScript to run only if it is being hosted on a trusted domain [Nosc2010]. The *BetterPrivacy* Firefox plugin tries to address the problem of supercookies by finding Flash cookies on the hard drive and regularly deleting them. Other tools include *Ghostery*, *Do Not Track Plus* and *AdBlock Plus*.

The Tracking Protection List (TPL) approach relies on a list, established by various organisations, that contains web addresses of misbehaving tracking sites.

Furthermore there are several privacy-enhancing tools that are not specific to third-party tracking but nevertheless provide some protection against it. Examples of such tools include private browsing modes of major browsers [Aggr2010] or anonymity networks [Ding2004].

Opt-out: Most tracking companies allow users to set opt-out cookies, and some tools such as *Beef Taco* make this process simpler. Many advertising networks interpret these cookies as opting out of receiving targeted advertisements, but still continue to track and profile the user. Most major browsers implement a DNT (Do Not Track) header that tells user-selected websites that they don't want to be tracked [DNT2011].

Privacy-by-Design and Privacy-Preserving systems: Privacy-by-design is often praised as an essential step towards better privacy protection: in a world where privacy is more and more jeopardised by new information and communication technologies, the growing view is that part of the remedy should come from the technologies themselves. On the technological front, privacy enhancing technologies (PETs) have been an active research topic in computer science for decades and a variety of techniques have been proposed (including anonymisers, identity management systems, privacy proxies, encryption mechanisms, anonymous credentials, etc.). However, privacy-by-design is more than the use of PETs: it relies on the idea that privacy requirements should be taken into account in the early stages of the design of a system and can have a potential impact on its overall architecture. In other words, privacy-by-design represents a paradigm shift: 'prevent rather than cure'. A number of general privacy principles have been proposed, such as the 'Fair Information Practice' principles back in 1974, including notice-and-choice, data integrity, and enforcement mechanisms.

A few behavioural advertising systems, such as *Adnostic*, *PrivAd* and *RePriv*, that consider privacy as one of the main design requirements, were proposed recently. The main objective of these schemes is to limit tracking, while still serving behavioural advertisements. *Privad* envisions a fully technical approach to non-tracking and private targeting [Guha2011]. The client builds a user profile and, according to this profile, requests relevant ads from the broker. A trusted party, the 'dealer,' anonymises the client to prevent the ad network from identifying the client. The anonymisation impacts performance and makes click-fraud harder to detect. In *Adnostic* the browser (via an add-on) continually updates a behavioural profile of the user based on browsing activity [Toub2010]. The ad network serves N (say, 10) ads instead of 1; the browser picks one based on the user's profile. Ad clicks are not considered private. Prohibition of tracking is contractual rather than technical. *RePriv* has the more general goal of enabling personalisation via interest profiling in the browser [Fred2011]. The applications are personalised search, site personalisation and ad targeting. Targeting is not done locally, however, and instead involves the browser sending the behavioural profile to the server.

8.2 Regulatory and Legislative approaches: What is done in the EU/ USA/ elsewhere?

8.2.1 European Union

The 2002 ePrivacy Directive, 2002/58/EC, mandated that websites must provide information about their data collection practices and must enable users to opt out of having information stored in their browser, except as 'strictly necessary' to provide service 'explicitly requested' by the user. In practice the directive has had little force; Member States have not taken any measures to enforce compliance, and in many cases they have treated browser cookie settings as adequate implementation [EU2011].

A 2009 amendment to the ePrivacy Directive, 2009/136/EC, replaces the opt-out rule with an opt-in consent rule [EU2010]. Member State implementations initially split. Some states have suggested existing browser settings would remain adequate, through the legal fiction that they convey 'implicit consent'. The majority view, as reaffirmed by the latest regulation proposal, is to require explicit, affirmative consent for each website [EU2012]. EU and state authorities have yet to enforce compliance with the rule.

8.2.2 United States

The Federal Trade Commission (FTC) is the leading federal regulatory agency for consumer protection. The FTC has narrowly circumscribed general statutory authority: It can only prevent 'unfair or deceptive' business practices, which the agency has largely interpreted to require violation of an express promise to consumers. The FTC can only levy monetary penalties against repeat offenders. In practice the FTC does, however, have a great degree of soft power: businesses are loath to endure the expense, burden, and optics of a federal law enforcement action. Signalling its heightened interest in the area, the FTC brought two enforcement actions related to third-party web tracking in 2011 [FTC2011a, FCT2011b].

State attorneys general have largely parallel authority in regulating third-party tracking. To date no attorney general's office has brought an enforcement action over tracking. Civil class action attorneys have attempted to raise a number of federal and state claims over third-party web tracking practices. In early litigation, several companies agreed to multi-million dollar settlements (e.g. Quantcast, over its use of Flash cookies [Mull2011a]). More recently the trend has been companies successfully having the cases against them dismissed [Mull2011b].

8.2.3 Online advertising self-regulation

The online advertising industry has largely harmonised self-regulatory efforts in the US (the Network Advertising Initiative, NAI, and the Digital Advertising Alliance, DAA) and the EU (the Interactive Advertising Bureau Europe, IAB Europe). All three programmes impose the same core requirements on behavioural advertising companies [NAI2008, DAA2009, IABE2011, DAA2011]:

1. They must provide users with information about their behavioural advertising practices.

2. They must allow users to opt out of behavioural advertising use of data. Note that this is a choice about one particular use of data; *collection and other uses of third-party tracking data are unaffected.*

Participation in self-regulation has fluctuated with regulatory attention [Gell2011]. At present most of the largest online advertising and analytics companies participate, and most of the smaller ones do not. Social networks and content providers are almost entirely absent. The Digital Advertising Alliance announced in late 2011 that it would attempt to expand its programme to all third parties and that it would broaden its consumer choice requirement to nearly all uses of third-party data for per-device (not per-user) personalisation [DAA2011]. Social networks and content providers have not yet signalled acceptance.

Researchers and civil society organisations have largely been critical of self-regulatory efforts for not providing choice over data collection and not imposing any meaningful punishments on companies that violate self-regulation.

8.2.4 Do Not Track (DNT)

Do Not Track is a technology and policy proposal that enables users to opt out of tracking by (all) websites they do not visit, including analytics services, advertising networks, and social platforms [DNT2011, Tene2011].

Technologically, the DNT mechanism is straightforward: the browser signals to websites the user's wish to opt out of tracking, specifically, via the 'DNT: 1' HTTP header. The header is sent out with every web request – this includes the page the user wishes to view, as well as each of the objects and scripts embedded within the page, including ads and trackers. It has been implemented, or is scheduled to be implemented in all major desktop browsers as of this writing. But in order for it to be meaningful, advertisers will have to respect the user's

preference not to be tracked. How would this be enforced? There is a spectrum of possibilities, ranging from self-regulation via the Network Advertising Initiative, to supervised self-regulation or 'co-regulation', to direct regulation. At the very least, by standardising the mechanism and meaning of opt-out, the DNT header promises a greatly simplified way for users to opt out compared to the current cookie mechanism [W3C]. Opt-out cookies are not robust, they are not supported by all ad networks, and are interpreted variously by those that do (no tracking vs. no behavioural advertising). The DNT header avoids these limitations and is also future-proof, in that a newly emergent ad network requires no new user action.

8.3 Educational approach

If consumers were better educated about the prevalence and consequences of online tracking, they would be able to make more informed decisions regarding their use of online technologies and services. Not only would they then better protect themselves with self-help tools, it would put competitive pressure on entities in the online tracking ecosystem and lead to a better functioning market. In this section we will consider what the avenues for consumer education are and how effective they are likely to be.

We can identify several types of existing consumer education efforts in the area of online tracking.

- *General advice about online privacy and raising awareness of the existence of the online tracking ecosystem*. The United States Federal Trade Commission offers tips on social networking safety[8] and online tracking.[9] ENISA has recently published a report on the privacy risks of cookies [Enis2011]. The Center for Democracy and Technology has published a guide to behavioural advertising, including third-party tracking and the privacy impact.[10] The European Safer Internet initiative promotes safer responsible use of the Internet to young people.[11]
- *Initiatives to inform consumers about self-defence tools*. Innumerable websites exhort consumers to periodically clear cookies and provide instructions for doing so. Stanford's donottrack.us and Mozilla's Do Not Track page both provide information on how to enable Do Not Track and what it will and will not do. Advocacy organisations such as the EFF frequently survey privacy technologies, including online tracking defences.[12] Naturally, vendors of defensive technologies are active in reaching out to the public to raise awareness of their products.[13]
- *Information about the data collection practices of specific companies and their products in the online tracking ecosystem*. Since the companies themselves are frequently the ones doing this, the category straddles the line between education and

[8] *http://www.ftc.gov/bcp/edu/pubs/consumer/tech/tec14.shtm*
[9] *http://onguardonline.gov/blog/curious-about-online-tracking-learn-about-cookies*
[10] *http://www.cdt.org/privacy/targeting*
[11] *http://www.saferinternet.org*
[12] *https://www.eff.org/wp/effs-top-12-ways-protect-your-online-privacy*
[13] *See for example http://www.prweb.com/releases/2011/10/prweb8883027.htm by the author of PrivacyChoice.*

transparency. Google offers an information page on advertising and privacy.[14] The media play a very important role in this type of education, frequently in the form of exposés of privacy violations by companies. *The Wall Street Journal*'s 'What They Know' series is the most well-known instance.[15] Recently, academic researchers have become more active in studying and exposing privacy violations in online tracking.

In light of the discussion above, there are a variety of entities with an interest in consumer education: governmental agencies, civil liberties and consumer advocacy organisations, the media, academics, companies involved in online tracking and vendors of privacy tools including web browsers. Not all entities engage in all categories of education, but each category is represented by multiple types of organisations.

As a closing remark for this section on 'education' it is also worth noting that numerous experimental results in the area of privacy research indicate that the prime interests for the majority of users of online services, are:

- Convenience / ease of use, and
- Cost of service (with a preference for 'free' offers);

Obviously both of these requirements imply the necessity for users to give away to the service providers personal information that is monetised by the providers in exchange for the 'free' services offered.

[14] *http://www.google.com/privacy/ads/ Unfortunately, the company continues to obfuscate the distinction between online tracking and behavioural advertising.*

[15] *http://online.wsj.com/public/page/what-they-know-digital-privacy.html*

9 Recommendations

While there are a few protective measures as discussed in the previous section, the current situation is not satisfactory from the point of view of consumers. We discuss some recommendations aimed primarily at regulators that, we believe, could help improve user privacy.

In this section we also take into account that the EC has recently published its proposal for a reform of the data protection rules [EU2012]. The Commission's proposal has now been passed on to the European Parliament and EU Member States (meeting in the Council of Ministers) for discussion, while the Regulation will be enforceable in all Member States two years after it has been adopted.[16]

1. Focus on tracking, not Online Behavioural Advertising (OBA)

Much of the debate today focuses on OBA instead of tracking. For example, some advertising companies interpret DNT as an opt-out of targeted behavioural ads, linking DNT to the industry self-regulatory programme. Tracking is the problem – not behavioural advertising. DNT should be interpreted as a request for not being tracked by third parties, either directly or with the help of first parties.

Recommendation aimed towards: EU policymakers (i.e. EC, EP), industry.

2. Elicit more meaningful privacy policies

Although the notice and choice paradigm, typically implemented via privacy policies, is often presented as a solution to privacy, it has serious limitations. First, privacy policies are usually long and complex to understand, and most users simply ignore them.[17] Second, the user's choice is typically binary – whether or not to use the product or service. Finally, not many users have the technical knowledge to fully understand the implications of consenting to behavioural tracking.

Despite these problems, privacy policies have an important role to play because they force companies to commit to their practices. Several ideas have been proposed for making notices more meaningful and comprehensible to consumers, such as 'visceral notice' [Calo2012]. Regulators have three roles to play: ensuring that companies specify privacy policies, incentivising companies to make these policies complete, concrete and meaningful, and ensuring compliance with stated policies. For example, the California Department of Justice has recently been taking steps to ensure that all mobile applications provide privacy policies.

Recommendation aimed towards: EU policymakers (i.e. EC, EP), industry.

---

[16] *Member States will also have a period of two years to transpose the provisions in the Directive into national law.*

[17] *It would take the average consumer more than 300 hours to read the privacy policies at the websites they visit each year, according to the high-end estimates of a 2008 study [Temp2012].*

3. Develop easy-to-use tools for transparency and control

As shown in [McDo2010] most users are unfamiliar with behavioural tracking and advertising. A large proportion of users are not even aware that they are being tracked and profiled while surfing the web, and that their profiles are used to deliver targeted ads. While periodically clearing cookies is one of the simplest (partial) counter-measures, only a minority of users understand what a cookie is, what they are used for and how to clear them. Users should be aware of how their data is being used/processed and what the potential dangers are. There is a need for enhanced *transparency* to help individuals to understand how their personal data (and, ideally, any data that can be used in a processing with potential effects on them) is collected, managed, and transferred. TETs (Transparency Enhancing Technologies) are critical given that information flows are growing dramatically and the data mining and inference techniques are becoming more and more powerful.

One initiative that goes in this direction is the icon-based program developed by a group of advertising associations [AOIP2012]. This program includes the use of an 'Advertising Option Icon' that marketers can place near their ads or on the Web pages that collect data that is used for behavioural targeting. Users who click on the icon see an explanation of why they are seeing a particular ad and are able to opt out of being tracked. Unfortunately this solution is not very good (icons are often very small and hard to see, and confusing) [Leon2011]. Users have difficulty distinguishing between tracking companies. Furthermore, the list of advertising companies and the technologies for tracking are changing constantly, making it difficult for tool providers, and users, to keep up.

On a more optimistic note, various entities like browser vendors, privacy advocates and the press have made vigorous efforts to develop more usable transparency tools. The Collusion Firefox add-on by Mozilla shows, in real time, all the third parties that are tracking the user across the Web. It generates a network of interactions between companies and trackers [Collu2012]. The WSJ Data Transparency Weekend[18] has provided funding and support for the development of some such tools; however, much more funding is necessary and can go a long way.

In light of the above, a possible first step could be for the European Commission (possibly in collaboration with European Agencies like ENISA) to launch an awareness campaign informing users how their data is being used/processed and what the potential dangers are.

Recommendation aimed towards: European Commission and/or ENISA, industry.

4. Develop compliance and monitoring initiatives

Compliance: Privacy impact assessment (PIA), and possibly privacy certification, should be promoted.[19] A PIA is a process used to determine how a service or application affects user

[18] *The Wall Street Journal Data Transparency Weekend, http://datatransparency.wsj.com/*

[19] *These compliance mechanisms are relatively new. It is still too early to tell whether they are an effective tool for the policymakers in the area of privacy.*

privacy. PIAs promote transparency and accountability, and contribute to public confidence in the way the service or application manages personal information and tracks users.

Monitoring/Detection violations: Opt-out or notice-based solutions are only effective if companies follow the rules and respect users' requests not to be tracked, as well as their own promises. While most of the large players do comply with self-regulatory standards, it has been shown that many smaller players do not [Koma2011].

The burgeoning 'web privacy measurement' community is a key part of the picture. Services and applications should therefore be scrutinised for data leakage/tracking, and the results should be made public. The UC Berkeley Web Privacy Census is an important step in this direction.[20] The media play a very important role in this information diffusion, as the *Wall Street Journal* is doing via its 'What They Know' series. It is important to note that since some types of tracking might be permitted, the tools in question are merely aids to determine when a further investigation is warranted.

There are a variety of passive ('fingerprinting') and active ('tagging') techniques to track users. Tagging is trivially detectable, since it requires modifying the state of the browser. As for fingerprinting, everything except for IP address and the user-agent string requires extra API calls and network activity that is in principle detectable. In summary, some crude tracking methods might be able to pass under the radar, while the finer-grained and more reliable methods are detectable. Detection of impermissible behavioural advertising is significantly easier. Intuitively, two users with DNT enabled should see roughly the same distribution of advertisements on the same web page, no matter how different their browsing history. In a single page view, there could be differences due to fluctuating inventories, testing, and randomness, but in the aggregate, two DNT users should see the same ads [Bale2012].

Enforcement: Solutions to block misbehaving players or force them to comply with the rules and legislation should be developed. One tricky issue for enforcement agencies is the issue of offshore tracking companies. A proposed solution is to disallow first parties from doing business with non-compliant third parties.

Recommendation aimed towards: EU policymakers (i.e. EC, EP).

5. Develop anti-tracking initiatives for Mobile Apps

Third-party tracking is proliferating on mobile platforms/smartphones [Thur2010]. However current browser-based solutions, such as DNT or TPL, have not yet been effectively adapted to mobile platforms – much of the third-party tracking on mobile devices happens in mobile applications, outside the context of a traditional browser. The tracking mechanisms are often embedded in applications, or rather in the advertising libraries they use [Grace2012]. Consequently, there is no way a user can express that he does not want to be tracked without uninstalling the applications. Solutions adapted to mobile platforms should be developed.

[20] *http://www.law.berkeley.edu/privacycensus.htm*

Recommendation aimed towards: European Commission and in particular EU-funded R&D programmes; industry.

6. Promote privacy-by-design

Although a few privacy-preserving alternatives to tracking have been proposed, as described in section 8.1, these solutions have seen little or no adoption. Indeed, we are not even aware of any serious pilot studies of such technologies: a necessary precursor to deployment. From the point of view of ad companies, social networks and other companies that collect data, privacy-preserving solutions come at a convenience (and possibly performance) cost, with no direct benefit. Pressure from privacy advocates has so far proved to be quite inadequate; regulation has an important role to play in incentivising companies to adopt privacy-preserving solutions. More broadly, the burden of enforcing online privacy should be shifted to businesses. This will push companies to integrate privacy into their products and processes, instead of disclaiming liability for privacy in legal notices.

Recommendation aimed towards: European Commission and in particular EU-funded R&D programmes; industry.

10 References

[Acqu2011] A. Acquisti, R. Gross and F. Stutzman, 'Faces of Facebook: Privacy in the Age of Augmented Reality', BlackHat Las Vegas, August 4, 2011, available at: http://www.heinz.cmu.edu/~acquisti/face-recognition-study-FAQ/, last visited September 2012.

[Aggr2010] G. Aggrawal, E. Bursztein, C. Jackson, and D. Boneh, 'An analysis of private browsing modes in modern browsers', in Proceedings of 19th Usenix Security Symposium, 2010.

[AOIP2012] Advertising Option Icon program. 2012. http://naiblog.org/tag/advertising-option-icon/, last visited September 2012.

[Ashk2009] Ashkan, S., S. Canty, M. Quentin, T. Lauren, and J. Chris, 'Flash cookies and privacy. Technical report', University of California, Berkeley, available at: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862, last visited September 2012.

[Bale2012] Rebecca Balebako, Pedro Leon, Richard Shay, Blase U, and Lorrie Cranor, 'Measuring the Effectiveness of Privacy Tools for Limiting Behavioral Advertising', Web 2.0 Security and Privacy Conference 2012.

[Boult2010] Boulton, C., 'Google CEO Schmidt Pitches Autonomous Search, Flirts with AI', available at: http://www.eweek.com/c/a/Search-Engines/Google-CEO-Schmidt-Pitches-Autonomous-Search-Flirts-with-AI-259984/1/, last visited September 2012.

[Bwee2007] 'So many ads, so few clicks', Business Week, November, 2007, available at: http://www.businessweek.com/magazine/content/07_46/b4058053.htm, last visited September 2012.

[Calo2012] R. Calo, 'Against Notice Skepticism', Notre Dame Law Review 1027, 2012.

[Cast2010] C. Castelluccia, E. De Cristofaro, and D. Perito, 'Private information disclosure from web searches', in Proceedings of the 2010 Privacy Enhancing Technologies Symposium (PETS), Berlin, Germany, 2010.

[Cast2012] Castelluccia, C., D. Kaafar and D.M Tran, 'Betrayed by Your Ads', in Proceedings of the 2012 Privacy Enhancing Technologies Symposium (PETS), Vigo, Spain, 2012.

[Chab2012a] A. Chaabane, G. Acs, M. A. Kaafar, 'You Are What You Like! Information leakage through users' Interests', The Network & Distributed System Security Symposium (NDSS), San Diego, 2012.

[Chab2012b] A. Chaabane, M. A. Kaafar, R. Borelli, 'Big Friend is Watching You: Analyzing online social networks tracking capabilities', in Workshop on Online Social Networks (WOSN'12), Helsinki, 2012.

[Collu2012] Mozilla Collusion add-on, available at: http://www.mozilla.org/en-US/collusion/, last visited September 2012.

[DAA2009] Digital Advertising Alliance, 'Self-regulatory principles for online behavioral advertising', July 2009, available at: http://aboutads.info/, last visited September 2012.

[DAA2011] Digital Advertising Alliance, 'Self-regulatory principles for multi-site data', November 2011, available at: http://aboutads.info/, last visited September 2012.

[Ding2004] R. Dingledine, N. Mathewson, and P. Syverson. 'Tor: The second-generation onion router', in Usenix security symposium, 2004.

[Dixo2011] Dixon, P. 'Consumer tips: How to opt-out of cookies that track you', available at: http://www.worldprivacyforum.org/cookieoptout.html, last visited September 2012.

[DNT2011] Do Not Track-Universal Tracking Opt Out, available at: http://donottrack.us/, last visited 2012.

[Ecke2009] Peter Eckerley, 'How Online Tracking Companies Know Most of What You Do Online', available at: https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks, last visited September 2012.

[Ecke2010] Eckerley, P., 'How unique is your web browser?', in Proceedings of the 2010 Privacy Enhancing Technologies Symposium (PETS), Berlin, Germany 2010.

[Enis2011] R. Tirtea, C. Castelluccia and D. Ikonomou (ENISA), 'Bittersweet cookies. Some security and privacy considerations', available at: http://www.enisa.europa.eu/act/it/library/pp/cookies/

[EU2010] Article 29 Data Protection Working Party, (2010, June) Opinion 2/2010 on online behavioural advertising, available at: http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf, last visited September 2012.

[EU2011] Article 29 Data Protection Working Party, (2011, August) Letter to the online advertising industry (OBA) Industry regarding the self-regulatory Framework, available at: http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2011/20110803_letter_to_oba_annexes.pdf, last visited September 2012.

[EU2012] European Commission, 'Commission proposes a comprehensive reform of the data protection rules', available at: http://ec.europa.eu/justice/newsroom/data-protection/news/120125_en.htm, last visited September 2012.

[FTC2011a] Federal Trade Commission (2011, March) 'FTC puts an end to tactics of online advertising company that deceived consumers who wanted to 'opt out' from targeted ads', available at: http://ftc.gov/opa/2011/03/chitika.shtm, last visited September 2012.

[FCT2011b] Federal Trade Commission (2011, November) 'Online advertiser settles FTC charges ScanScout deceptively used Flash cookies to track consumers online', available at: http://www.ftc.gov/opa/2011/11/scanscout.shtm, last visited September 2012.

[Fred2011] Matthew Fredrikson and Benjamin Livshits. 'RePriv: Re-envisioning in-browser privacy', in IEEE Symposium on Security and Privacy, May 2011.

[Gell2011] R. Gellman and P. Dixon, 'Many failures: A brief history of privacy self-regulation in the United States', October 2011, available at: http://worldprivacyforum.org/pdf/WPFselfregulationhistory.pdf, last visited September 2012.

[GoAn] Google Analytics, Enterprise-class web analytics, available at: http://www.google.com/analytics/, last visited September 2012.

[Grace2012] Michael Grace et al., 'Unsafe Exposure Analysis of Mobile In-App Advertisements', ACM Wisec 2012.

[Gree2008] Greene, K. (March/April2008) 'Reality mining', MIT Tech. Review, available at: http://www.technologyreview.com/read_article.aspx?id=20247&ch=specialsections&sc=emerging08&pg=1, last visited September 2012.

[Guha2011] S. Guha, B. Cheng and P. Francis. 'Privad: Practical Privacy in Online Advertising', in Proceedings of the 8th Symposium on Networked Systems Design and Implementation (NSDI), Boston, MA, Mar 2011.

[IABE2011] Interactive Advertising Bureau Europe, 'IAB Europe EU framework for online behavioural advertising', July 2011, available at: http://www.iabeurope.eu/media/55448/iab%20europe%20report%20july_28.pdf, last visited September 2012.

[Kamk2010] Kamkar, S. (October 2010). 'Evercookie – never forget', available at: http://samy.pl/evercookie/, last visited September 2012.

[Koma2011] Saranga Komanduri, Richard Shay, Greg Norcie, Blase Ur, Lorrie Faith Cranor, 'AdChoices? Compliance with Online Behavioral Advertising Notice and Choice Requirements', CMU TR CMU-CyLab-11-005, March 2011.

[Krish2009a] Krishnamurthy, B. and C. Wills, 'On the leakage of personally identifiable information via online social networks', in WOSN '09: The second workshop on Online social networks.

[Krish2009b] Krishnamurthy, B. and C. Wills, 'Privacy diffusion on the web: a longitudinal perspective', in WWW '09: Proceedings of the 18th international conference on World wide web. ACM.

[Krish2009c] Krishnamurthy, B. and C. Wills, 'Privacy diffusion on the web: a longitudinal perspective' (updated graphs), available at: http://www.ftc.gov/os/comments/privacyroundtable/544506-00009.pdf, last visited September 2012.

[Krish2011] Balachander Krishnamurthy, Konstantin Naryshkin and Craig Wills, 'Privacy leakage vs. Protection measures: the growing disconnect', Web 2.0 Security and Privacy Workshop, May 2011.

[Leon2011] Pedro G. Leon, Blase Ur, Rebecca Balebako, Lorrie Faith Cranor, Richard Shay, and Yang Wang, 'Why Johnny Can't Opt Out: A Usability Evaluation of Tools to Limit Online Behavioral Advertising', CMU technical report CMU-CyLab-11-017, October 31, 2011.

[Macm2009] Macmanus, M., A Guide to Recommender Systems, Jan. 2009, available at: http://www.readwriteweb.com/archives/recommender_systems.php, last visited September 2012.

[Maye2012] J. Mayer and J. Mitchell, 'Third-Party Web Tracking: Policy and Technology', IEEE Security&Privacy, San Francisco, May 2012.

[McDo2010] McDonald, A. M., and Cranor, L. F. 'Americans' Attitudes about Internet Behavioral Advertising Practices', Proceedings of the 9th Workshop on Privacy in the Electronic Society (WPES) October 4, 2010.

[Mcki2008] McKinley, K., 'Cleaning up after cookies. Technical report', iSEC PARTNERS, available at: https://www.isecpartners.com/files/iSEC_Cleaning_Up_After_Cookies.pdf, last visited September 2012.

[Mull2011a] J. Mullen, 'Judge approves $2.4 million Quantcast privacy settlement', paidContent, available at: http://paidcontent.org/2011/06/15/419-judge-approves-2-4-million-quantcast-privacy-settlement/, last visited September 2012.

[Mull2011b], 'Second privacy law suit over cookies' falls apart', paidContent, available at: http://paidcontent.org/2011/08/19/419-privacy-lawsuits-over-flash-cookies-falling-apart/, last visited September 2012.

[NAI2008] Network Advertising Initiative, '2008 NAI principles', available at: http://networkadvertising.org/, last visited September 2012.

[Nara2008] Arvind Narayanan, Vitaly Shmatikov. 'Robust de-anonymization of large sparse datasets', IEEE S&P '08.

[Nara2011a] A. Narayanan. 'There is no such thing as anonymous online tracking', available at: http://cyberlaw.stanford.edu/node/6701, last visited September 2012.

[Nara2011b] Arvind Narayanan, 'Price Discrimination is All Around You', http://33bits.org/2011/06/02/price-discrimination-is-all-around-you/, last visited September 2012.

[Nara2011c] Arvind Narayan, 'The Linkability of Usernames: a Step Toward "Uber-profiles"', blog posts, Feb. 2011. http://33bits.org/2011/02/16/usernames-linkability-uber-profiles/, last visited September 2012.

[Nosc2010] NoScript Firefox extension, 2010, available at: http://noscript.net/, last visited September 2012.

[NYT2012] C. Duhigg, 'How Companies Learn Your Secrets', The New York Times, February 2012, available at: http://www.nytimes.com/2012/02/19/magazine/shopping-habits.html?pagewanted=all, last visited September 2012.

[Pari2011] Eli Pariser, The Filter Bubble: What the Internet is Hiding from You, Penguin Press, March 2011.

[Peri2011] Daniele Perito, Claude Castelluccia, Mohamed Ali Kaafar, Pere Manils. 'How Unique and Traceable are Usernames?', 11th Privacy Enhancing Technologies Symposium (PETS 2011), Waterloo, CA, 2011.

[Raph2011] Raphael, J. R., 'Apple vs. Android location tracking: Time for some truth', http://blogs.computerworld.com/18190/apple_android_location_tracking, last visited September 2012.

[Roes2012] Franziska Roesner, Tadayoshi Kohno, and David Wetherall, 'Detecting and Defending Against Third-Party Tracking on the Web', at 9th USENIX Symposium on Networked Systems Design and Implementation (NSDI 2012), San Jose, CA, April 2012.

[Scho2009] Schoen, S., 'New Cookie Technologies: Harder to See and Remove, Widely Used to Track You', available at: http://www.eff.org/deeplinks/2009/09/new-cookie-technologies-harder-see-and-remove-wide, last visited September 2009.

[Solo2011] Solove, Daniel J., 'Nothing to Hide: The False Tradeoff between Privacy and Security', May 1, 2011. GWU Law School Public Law Research Paper No. 571.

[Temp2012] J. Temple, 'Why Privacy Policies don't work', San Francisco Chronicle, Jan. 2012. http://www.sfgate.com/cgi-bin/article.cgi?f=/c/a/2012/01/28/BU5N1MUKGO.DTL#ixzz1lWUW5RX8 , last visited September 2012.

[Tene2011] Omer Tene and Jules Polonetsky, 'To Track or "Do Not Track": Advancing transparency and individual control in online behavioral Advertising', Minnesota Journal of Law, Science & Technology, Volume 13, Issue 1, Winter 2012.

[Thur2010] S. Thurm and Y. Kane, 'Your Apps Are Watching You', The Wall Street Journal, Dec. 2010. http://online.wsj.com/article/SB10001424052748704694004576020083703574602.html, last visited September 2012.

[Toub2010] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum, and Solon Barocas, 'Adnostic: Privacy preserving targeted advertising', NDSS, San Diego, USA, 2010.

[W3C] Tracking Protection Working Group. W3C. http://www.w3.org/2011/tracking-protection/, last visited September 2012.

[WSJ] 'What They Know about You!', Wall Street Journal, available at: http://online.wsj.com/public/page/what-they-know-digital-privacy.html, last visited September 2012.

[Yen2012] Ting-Fang Yen, Yinglian Xie, Fang Yu, Roger P. Yu, Martin Abadi, 'Host Fingerprinting and Tracking on the Web: Privacy and Security Implications', in Proceedings of the 19th Annual Network & Distributed System Security Symposium (February 2012).

P.O. Box 1309, 71001 Heraklion, Greece
www.enisa.europa.eu