

Chapter 2

Behavioural Tracking on the Internet: A Technical Perspective

Claude Castelluccia

2.1 Behavioural Tracking

2.1.1 *The Context: Behavioural Profiling*

The concept of *Behavioural Profiling* (also known as “targeting”) consists of collecting and analysing several events, each attributable to a single originating entity, in order to gain information relating to the originating entity. It consists of, in other words, transforming data into knowledge (Hildebrandt 2006). Behavioural profiling involves collecting data (recording, storing and tracking) and searching it for identifying patterns (with the help of data mining algorithms). The data collection phase is often referred to as *Behavioural Tracking*.

An example of behavioural targeting scenario is provided in Dwyer (2009). A consumer shops online for an airline ticket to New York City. He searches for flights, but does not make any purchase. He subsequently visits the web site of the local newspapers that displays adds offering tickets to New York. While no Personally Identifiable Information (PII) might have been collected, his interest in airline tickets has been noted.

2.1.2 *Motivations: Why are We Being Tracked and Profiled?*

Profiles are very valuable for many companies in customising their services to suit their customers, in order to increase revenues. The clear intent of behavioural targeting is to track users over time and build profiles of their interests, characteristics (such as gender, age and ethnicity) and shopping activities. For example, advertising or publishing companies use behavioural targeting to display advertisements that

C. Castelluccia (✉)
INRIA Rhone-Alpes, Grenoble, France
e-mail: claudc.castelluccia@inria.fr

closely reflect the users interests. Online advertising systems are typically composed of three main entities: the *advertiser*, the *publisher* and the *ad network*. The advertiser is the entity, for example a car manufacturer or a hotel, which wishes to advertise a product or service. The publisher is the entity, such as an online newspaper company, which owns one or several web sites 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 the user clicks on an advertisement, the ad network collects payment from the corresponding advertiser. There is, therefore, a strong incentive for the ad network to generate very accurate and complete profiles in order to maximise profit. E-commerce sites also use behavioural tracking to recommend products that are likely to be of interest to users. For example, Amazon recommends products to online users based on the individuals past behaviour (personalised recommendation), on the past behaviour of similar users (social recommendation) and, of course, on the searched items (item recommendation; Macmanus 2009).

2.1.3 *Tracking and Privacy*

It can be argued that the customisations resulting from profiling are also beneficial to the users that only receive information relevant to their interest. However, it creates serious privacy concerns since it allows some companies or institutions to gather and concentrate a huge amount of information about their customers, and about Internet users in general.

The danger is to move into a surveillance society or Internet, where all our online or physical activities are recorded and correlated. Some companies offer various services that gather different types of information from users. The combination and concentration of all this information provides a powerful tool to accurately profile users. For example, Google is one of the main third-party aggregators and tracks users across most web sites (Krishnamurthy and Willis 2009b). In addition, it also runs the most popular search engine and, as such, stores web histories of most users (i.e. their search requests), their map searches (i.e. their requests to the Google map service), their images, etc. (Castelluccia et al. 2010). Web searches have been shown to often be sensitive (Conti and Sobiesk 2007). It has actually been demonstrated that it is quite trivial to derive the identity of a user from his web history (Barbaro and Zeller 2006). Map requests also leak a lot of information, such as the user's home address or his favourite places. Finally, Google runs one of the most popular email systems, *gmail*, and has, therefore, access to emails of millions of users. By combining these different types of information coming from different sources, Google is able to build very accurate profiles of their users. As argued in Hildebrandt (2006), "profiling shifts the balance of power between those that can afford profiling (mostly large organisations) and those that are being profiled (mostly individual citizens), because the profilers have a certain type of knowledge to which those profiled have no effective access."

The advent of *ubiquitous advertising*, which can be seen as the application of *computational advertising*¹ to smart phones, will provide even more sources of profiling information (Krumm 2010). With ubiquitous advertising, advertisements will not only be personalised to users' online profiles, but also to their physical profiles. Advertisements will be customised to users' locations, physical or intellectual activities, interactions and possibly moods. Since, as opposed to a regular computer, a mobile device is usually owned by a single person, more detailed and accurate profiles can be derived from his uses. It is also foreseen that, in the future, sensors on phones will be able to infer users' food habits and preferences (Krumm 2010). These new developments create serious privacy issues that need be studied more carefully (Clegg 2007).

The rest of this chapter considers three of the most popular Internet services, namely the web, location-based services (LBS) and online social networks (OSN). It presents for each of them existing tracking mechanisms. Note that we do not cover the profiling part, which consists on transforming collected data into knowledge. Furthermore, it focusses on technological issues, and eludes legal or policy aspects.

2.2 Web Tracking

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 includes 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 using techniques such as *cookies*, *Javascripts* or *supercookies* (McKinley 2008).

Cookies A cookie is a piece of text stored by a user's web browser and associated to a HTTP request. A cookie 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 preferences. They are set by a service when a user logs in, and are erased when the user logs out. Persistent cookies are often used as 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 web site and can, therefore, be used by web sites to track users across visits. Persistent cookies raise serious privacy concerns. In the rest of the document, the term cookie refers to persistent cookie, unless explicitly stated.

Cookies are sent only to the web sites that set them or to servers in the same Internet domain. However, a Web page may contain images, links, web bugs (1 × 1 pixel GIF images), HTML IFrames, Javascript or other components stored on servers

¹ Computational advertising is a new scientific sub-discipline whose main challenge is to find the best ad to present to a user engaged in a given context (Broder and Josifovski 2010).

in other domains. Cookies that are set during retrieval of these components are called *third-party cookies*,² 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 user's presumed preferences. Third-party tracking raises serious privacy concerns, which are not hypothetical but real. The increasing presence and tracking of third-party sites used for advertising and analytics has been demonstrated in a study (Krishnamurthy and Willis 2009b, 2009c). This study showed that the penetration of the top 10 third-parties growing from 40% in 2005 to 70% in 2008, and to over 70% in September 2009. Another study shows that not only these third-parties are increasing their tracking of users, but also they can now link these traces with identifiers and personal information via OSN (Krishnamurthy and Willis 2009a). In Dwyer (2009), a behavioural targeting study was performed on the *levis.com* site, the e-commerce site for the clothing line. The results show that the web site contains a total of nine tracking tags that link to eight third-party companies.³

Javascripts Many web sites contain executable Javascript files that are down-loaded by visiting users. These files, in addition to their computations, sometimes update first-party cookies and send information back to the servers. Javascripts 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, the tracking sites have all the regular information available in a typical HTTP request: sender's IP address, user-agent software information, current and previous URL (via Referer header), email address (from 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 (Dixon 2011). Therefore, this is a big impetus in the Internet tracking industry to discover and deploy more robust tracking mechanisms, often referred to as *Supercookies* (McKinley 2008). One of the most prominent supercookies is the so-called "Flash cookie", a type of cookie maintained by the Adobe Flash plug-in on behalf of Flash applications embedded in web pages (Schoen 2009). Since these cookie files are stored outside of the browser's control, web browsers do not directly allow users to control them. 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 traditionally 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

² Some sites included JavaScript code and third-party cookies from more than ten different tracking domains (Eckersley 2009).

³ The largest third-party Ad-network companies include *Advertising.com*, *Tacoda*, *DoubleClick* and *Omniure*. Most of these networks are owned by *Google*, *Yahoo*, *AOL* or *Microsoft*. Since Ad-networks are typically partnered with many publishers, they can track users across several publishers and build these users' browsing profiles.

user's HTTP cookie policies and privacy preferences. For example, it was found that some sites use HTTP and Flash cookies that contain redundant information (Ashkan et al. 2009). 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 as illustrated recent *evercookies* (Kamkar 2010). This new type of cookie identifies a client even when standard cookies, Flash cookies, and others, have been removed. This is accomplished by storing the cookie material in several types of storage mechanisms that are available on the local browser.

Browser fingerprinting A recent study showed that browsers can be identified to a high degree of accuracy without cookies or other tracking technologies (Eckersley 2010). Every Web browser provides enough unique information (User Agent, fonts, screen resolution, ...) to tell one from another. The study shows that a browser fingerprint is unique enough that it can, on the average, identify a browser among a set of 286.777 other browsers. Browser fingerprinting is a powerful tool for tracking users. It should be considered alongside with IP addresses, cookies and supercookies as far as user traceability is concerned.

2.3 Location Tracking

2.3.1 Location Privacy

More and more systems and applications record user's locations and movements in public places. These systems provide very useful and appreciated services, and have come to be regarded as almost essential and inevitable. For example, RFID cards allow users to open doors or pay their transportation ticket; GPS systems help users to navigate and find their ways. Some services tell users where their friends are, or provide personalised services (such as indicating the closest restaurant or hotel). Some wireless parking meters send users a text message when their time is running out (Blumberg and Eckersley 2009). 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 (Raphael 2011).

Location privacy is often defined as *the ability of an individual to move in public space with the expectation that their location will not be systematically and secretly recorded for later use*. Location tracking is not a new phenomenon, but new technologies (wireless networks, digital cameras, etc.) make it cheaper and easier to perform. It is this transformation to a world where location is collected pervasively, silently and cheaply that is worrisome (Blumberg and Eckersley 2009).

2.3.2 Location-based Services

Already today, worldwide, hundreds of millions of people permanently hold at least one mobile phone. It is predicted that smartphones will surpass PC sales within two

years (Boulton 2010). These mobile phones have increasing computational capacities and are equipped with multiple sensors like microphones, cameras, GPS, accelerometers, etc. As geolocated systems, they already enable individuals and communities to collect and share various kinds of data. Urban sensing is a new sensing paradigm leveraging users as part of a sensing infrastructure (Campbell et al. 2006). In the near future, several urban sensing applications are likely to appear, which will provide extra information about users (Miluzzo et al. 2008). Most users are unaware of the extra information that is collected about them beyond requested data, especially in case of participatory sensing. For example, a picture taken by a user may reveal additional contextual information inferred from the background or the style of any associated text. A recent study showed that most people are unaware of the fact that the photos and videos taken with their smart phones or cameras contain geolocation information (Friedland and Sommer 2010). This information can be used to localise them while they are travelling, or even reveal their home address. This may be considered as a potential source of information leakage and may lead to a privacy breach if used for location tracking or in conjunction with data retrieved from OSN. The risk becomes higher as the border between OSN and LBS becomes fuzzier. For instance, OSN such as *FourSquare*⁴ and *Gowalla*⁵ are designed to encourage users to share their geolocated data. Information posted on social applications such as *Twitter*⁶ can be used to infer whether or not an individual is at home.⁷ Other applications, such as *GoogleLatitude*,⁸ 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 exploited by potential adversaries, such as the free geographic data provided by *Google Maps*,⁹ *Yahoo! Maps*¹⁰ and *Google Earth*.¹¹

The W3C geolocation API, which is supported in the Firefox, Opera and Chrome browsers and in Internet Explorer via a plug-in, allows web sites to request geographical information for the clients device. With the approval of the user, the browser sends information like the clients 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 web site with the users 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 within

⁴ <http://foursquare.com/>.

⁵ <http://gowalla.com/>.

⁶ <http://twitter.com/>.

⁷ <http://pleaserobme.com/>.

⁸ <http://www.google.com/latitude/>.

⁹ <http://maps.google.com/>.

¹⁰ <http://maps.yahoo.com/>.

¹¹ <http://earth.google.com/>.

the range of an 801.11 wireless-base station (i.e. tens of meters). Furthermore, a growing number of sites now provide public APIs to their geolocalised content. For example, *Flickr*, *YouTube* and *Twitter* allow queries for results originating at a certain location. PicFog, for example, uses one of these APIs to provide real-time location-based search of images posted on Twitter. As shown in Friedland and Sommer (2010), these APIs can also be used to identify the current location of a user while he or she is away from home.

The emergence of *Reality Mining* raises even more privacy concerns (Greene 2008). As Greene (2008) explained, reality mining infers human relationship and behaviour from information collected by cellphones. This information include data collected by cellphone sensors, such as location or physical activity, and data recorded by phones themselves, such as call duration and 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 (Boulton 2010). 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.

2.4 Social Network Tracking

2.4.1 Online Social Networks

OSN have gained an immense popularity in recent years. Social-based services such as *Facebook*,¹² *Twitter*, *MySpace*¹³ and *Orkut*,¹⁴ just to name a few, allow millions of individuals to share some of their personal information with a multitude of other entities, such as their friends, companies or even the public at large. The common characteristic of these OSN is that users can make contacts and share easily personal information on a large scale. More specifically, people can meet old as well as new friends (*Facebook*, *MySpace*), find new jobs (*LinkedIn*¹⁵), or receive and provide recommendations (*Tribe*¹⁶). In a near future, many more complex services are likely to appear, which will tap on the power of the social connection and personal information provided by OSN.

¹² <http://facebook.com/>.

¹³ <http://www.myspace.com/>.

¹⁴ <http://www.orkut.com/>.

¹⁵ <http://www.linkedin.com/>.

¹⁶ <http://www.tribe.net/>.

As the primary objective of most of these services is to make individuals or groups visible, people need to share personal information to ensure some form of identifiability. Hence, most OSN encourage users to publish personal information, which may enable anyone accessing this information to infer further private information, thus causing a privacy breach. On top of that, the majority of users are not only willing but also pleased to disclose their personal information to as many users as possible and some OSN make this information public by default. Moreover, compared to traditional off-line, real-life, social networks, OSN are usually larger and contain more ties. For instance, people easily classify thousands of users as “friends”, or as “friends of friends”, when they probably would not qualify some of these users as friends in their real life. These facts inherently entail the question of trust and privacy in OSN.

Generally, average users do not have a clear idea of who accesses their private information, or what portion of it really needs to be accessed by applications. For instance, in Facebook, the terms of use of some applications clearly state that these applications can access any personal information put by the user, even though it may not be required. Although most sites provide coarse-grained privacy controls, the majority of users do not use this feature because they find it too complex (Gross et al. 2005). Moreover, these sites are permissive and allow anyone to access user’s profile data, which means that, by default, it is accessible by any other user in the network. In addition, it is difficult for an average user to know and control users or groups of users who can access his information and to limit this access without losing the benefits of the various features of OSN.

Another problem stems from the fact that while a user’s profile may be set to be inaccessible for other users, the friendship links and group affiliations often remain public. This public social information can leak further information about the private attributes of a profile. For instance, Zheleva and Ghetoor (2009) have shown that the structure of the social network and group information leak a surprisingly large amount of personal information. Moreover, even if a user makes some parts of his profile private, the person’s membership in a particular group remains publicly accessible from the group profile. Another study lead by MIT students, called the Gaydar project, has shown that it is possible to predict with a fairly high accuracy the sexual preferences of an individual. This is possible even if his profile is private, just by looking at the amount of gay friends it includes, compared with a person sampled randomly from the population (Johnson 2009).

Furthermore, much like traditional web sites, third-party aggregators track user activity pervasively on OSN (Krishnamurthy and Willis 2008). Third-party domains are then not only able to track the web sites that a user visits, but also the OSN sites that he connects to. In a follow-up work (Krishnamurthy and Willis 2009a), the same authors demonstrate that PII belonging to any user, such as name, gender or OSN unique ID, is also being directly leaked to these third-party servers via the OSN. This leakage happens via a combination of HTTP header information and cookies being sent to third-party aggregators. This result implies that third parties are not only able to view the surfing habit of some users, but are also able to associate the habits with a specific habit and potentially gather much more personal information. This ability to link information across web sites and OSN raises important privacy concerns.

2.4.2 Mobile Online Social Networks

Mobile Online Social Networks (MOSN) have recently grown in popularity. Mobile devices provide ubiquitous access to the web and naturally to social networks. There are typically two classes of mobile OSN: (1) traditional OSN (such as *Facebook*, *Twitter*) that have created content and access mechanisms tailored to mobile devices, and (2) new MOSN, such as *Foursquare* and *Loopts*,¹⁷ created to deal with the new mobile context. These new MOSN tend to customised their content to the location and the user's community (friends). For example, using the phone's self-location features, as well as information about the prior activities of the user's friends, some MOSN propose new places to explore or activities to try. Other MOSN allow a user to locate his friends that are currently in his or her vicinity. The predominant concepts of new MOSN are *presence* and *location* (Krishnamurthy and Willis 2010). Presence allows a user to know the current status of his or her friends. The indication of presence allows the expectation of a quick response. Location allows a user to locate his friends and obtain LBS, such as the closest restaurants or hotels. A recent study showed that most MOSN leak some kind of private information to users within the same MOSN, to users within other OSN via the interconnect features and, and more importantly, to third-party tracking sites. In many cases, data given out contained user's precise location, his gender or name, and even subject's unique social networking identifier, which could allow third-party sites to connect the records they keep of users' browsing behaviour with their profiles on the social networking sites.

The combination of location information, unique identifiers of devices, and traditional leakage of other personally identifiable information now give third-party aggregation sites the capacity to build a comprehensive and dynamic portrait of MOSN users.

2.5 Discussion

As illustrated in this report, users are being constantly tracked and profiled when using the Internet. This profiling will increase with the development of ubiquitous advertising and personalised services.

Unfortunately, there is no easy way to use modern, cookie- and JavaScript-dependent web sites and social networking sites and avoid tracking at the same time (Eckersley 2009). However, although not perfect (Aggrawal et al. 2010), private browsing mode of major browsers, that disable cookies, should be used when possible. Also, the popular Firefox *NoScript* extension should be considered. *NoScript* (2010) is a Firefox add-on which allows executable content such as JavaScript to run only if it is being hosted on a trusted domain. Finally, anonymisation networks, such as *TOR* (Dingledine et al. 2004), and network/web proxies that allow users to surf the Internet anonymously, mitigate some of the highlighted privacy issues.

¹⁷ <http://www.loopts.com/>.

As suggested in Schoen (2009), privacy-invasive marketing practices need greater scrutiny. More research is needed to reveal how the other kinds of cookies described in McKinley (2008) are also being used to track users. There is a lot of work to be done to bring these next-generation cookies even to the same level of visibility and control that users experience with regular HTTP cookies. Application and Browser developers should do more to let users control how they are being tracked. However, this is not an easy task since, as shown previously, some of these tracking cookies, such as the Flash ones, are stored outside of the browser. The *BetterPrivacy* Firefox plug-in tries to address this problem by finding Flash cookies on the hard drive and regularly deleting them.

In this context, it is going to be challenging to protect users' privacy. Some people argue that abstinence or withdrawal from the online world is the only method guaranteed to work (Conti 2009), or that users should lower their privacy expectation. According to Eric Schmidt, executive chairman of Google, it is possible to identify a person from 14 of his photos and then search the Web for more content about this user. Furthermore, he argues that, in the future, not only we will be able to identify a person but also predict, from his messaging and location, where that person is going to go (Kirkpatrick 2010).

Users should be given the ability to control access and distribution of their personal data. Once data is used without the knowledge or consent of the user, privacy is clearly compromised. Solving these privacy issues will be beneficial not only to users but also to service providers. In fact, as argued in Cleff (2007), users might react to this privacy fear by restricting the information they provide or by providing false information. This would have for effect to limit business, and to affect the validity of customer databases and profiles.

Users must also be able to choose what data is collected about them. They must keep the right to access, modify and delete them. Users should be explicitly informed about how they are being tracked, how their data is being sent/leaked out of their social network sites, by advertisers or others, and the corresponding destination. For example, users should need to acknowledge usage of their location on a per-application basis, or even, for some applications, each time location information is used. A simple, yet promising, approach is the *Do Not Track* (DNT) initiative. DNT gives users a way to opt out of behavioural tracking universally. In its simplest form, DNT is implemented as a HTTP header. This header contains a "Do-Not-Track" flag that indicates to web sites the user's wish to opt out of tracking. This extension is simple to implement in the web browser. As a matter of fact, there is already a Firefox add-on that implements such a header. However, this solution will only be effective if advertisers will respect the user's preference of not being tracked. As discussed in Narayanan (2010), there are several possibilities to enforce it, ranging from self-regulation via the Network Advertising Initiative, to supervised self-regulation or direct regulation.

Furthermore, more tools to help users making informed decisions about the publication of their data or their online activities should be developed. These tools should, for example, inform users whether the information to be published can potentially be combined with other data on the Internet to infer sensitive information

(Chew et al. 2008). *ReclaimPrivacy*¹⁸ is an example of such tools. *ReclaimPrivacy* is an open tool for scanning Facebook privacy settings and warn users about settings that might be unexpectedly public.

Finally, services and networks should follow the “privacy by design” concept (Le Métayer 2010). Privacy should be seen as a main design requirement, not as an add-on. For example, data collection should be minimal and only performed when necessary. Services should potentially be distributed and open-source to minimise data monitoring and collecting.¹⁹ They should request and use users’ identities only when strictly necessary. For example, most LBS request users to provide their identity before offering their services. This is required for accounting and billing purposes. However, the only thing that service operators actually need is an anonymous proof that the user is a registered subscriber (Blumberg and Eckersley 2009). This can be achieved, without revealing the user’s identity, by using existing cryptographic primitives (Zhong et al. 2007).

In summary, networks and services should be designed to limit unnecessary data collection and give individuals control over their data (Castelluccia and Kaafar 2009; Schneier 2009). Indeed as argued by Bruce Schneier (2009), Privacy is not something that appear naturally online, it must be deliberately architected. Privacy issues in behavioural profiling are complex and cannot be treated exclusively by technological means. There is a need for a true research approach that considers educational, policy, legal and technological aspects.

Acknowledgement The author would like to thank the members of the INRIA Planete group for discussions and for proofreading this chapter. He would also thank Levente Buttyan, Imad Aad, Aurelien Francillon, Bala Krishnamurthy, Emiliano De Cristofaro and many others for providing comments on this chapter. Finally, the author would like to thank ENISA and more particularly Rodica Tirtea who was at the origin of this work and chapter. This chapter was published as a section of the Privacy, Accountability and Trust Challenges and Opportunities report, published by ENISA (2011).

References

- Aggrawal, G., E. Bursztein, C. Jackson, and D. Boneh. 2010. An analysis of private browsing modes in modern browsers. Proceedings of 19th Usenix Security Symposium. Washington D.C., U.S.A.
- Ashkan, S., S. Canty, M. Quentin, T. Lauren, and J. Chris. 2009. *Flash cookies and privacy*. Technical report, University of California, Berkeley. http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1446862. Accessed in November 2010.
- Barbaro, M., and T. Zeller. 2006. A face is exposed for AOL searcher no. 4417749. *New York Times*, 9. August.
- Blumberg, A., and P. Eckersley. 2009. On locational privacy, and how to avoid losing it forever. <http://www.eff.org/wp/locational-privacy>. Accessed in November 2010.

¹⁸ <http://www.reclaimprivacy.org/>.

¹⁹ The Disapora project, see <http://www.joinindiaspora.com/>.

- Boulton, C. 2010. Google CEO Schmidt Pitches autonomous search, flirts with aI. <http://www.eweek.com/c/a/Search-Engines/Google-CEO-Schmidt-Pitches-Autonomous-Search-Flirts-with-AI-259984/1/>. Accessed in November 2010.
- Broder, A., and V. Josifovski. 2010. Introduction to computational advertising. <http://www.stanford.edu/class/msande239/>. Accessed in November 2010.
- Campbell, A. T., S. B. Eisenman, N. D. Lane, E. Miluzzo, and R. A. Peterson. 2006. People-centric urban sensing (invited paper). Proceedings of the Second ACM/IEEE International Conference on Wireless Internet. Boston, MA, U.S.A.
- Castelluccia, C., E. De Cristofaro, and D. Perito. 2010. Private information disclosure from web searches. Proceedings of the 2010 Privacy Enhancing Technologies Symposium (PETS). Berlin, Germany.
- Castelluccia, C., and D. Kaafar. 2009. Ocn: Owner-centric networking. In Future Internet Security and Trust (FIST) workshop. Seattle, WA, U.S.A.
- Chew, M., D. Balfanz, and B. Laurie. 2008. (under) mining privacy in social networks. Web 2.0 Security and Privacy workshop. Oakland, CA, U.S.A.
- Cleff, E. B. 2007. Privacy issues in mobile advertising. *International Review of Law, Computers & Technology* 21 (3): 225–236.
- Conti, G. 2009. *Googling security: How much does Google know about you?* Boston: Addison-Wesley.
- Conti, G., and E. Sobiesk. 2007. An honest man has nothing to fear: User perceptions on web-based information disclosure. Proceedings of the 3rd SOUPS' 07, New York, pp. 112–121.
- Dingledine, R., N. Mathewson, and P. Syverson. 2004. Tor: The second-generation onion router. Proceedings of Usenix security symposium. San Diego, CA, U.S.A.
- Dixon, P. 2011. Consumer tips: How to opt-out of cookies that track you. <http://www.worldprivacyforum.org/cookieoptout.html>. Accessed in July 2011.
- Dwyer, C. 2009. Behavioral targeting: A case study of consumer tracking on levis.com. Proceedings of Fifteen Americas Conference on Information Systems. San Francisco, CA, U.S.A.
- Eckersley, P. 2009. How online tracking companies know most of what you do online. <https://www.eff.org/deeplinks/2009/09/online-trackers-and-social-networks>. Accessed in November 2010.
- Eckersley, P. 2010. How unique is your web browser? Proceedings of the 2010 Privacy Enhancing Technologies Symposium (PETS). Berlin, Germany.
- ENISA. 2011. *Privacy, accountability and trust challenges and opportunities*. Technical report, ENISA.
- Friedland, G., and R. Sommer. 2010. Cybercasing the joint: On the privacy implication of geo-tagging. Usenix Workshop on Hot Topics in Security. Washington D.C., U.S.A.
- Greene, K. 2008. Reality mining. http://www.technologyreview.com/read_article.aspx?id=20247&ch=specialsections&sc=emerging08&pg=1. Accessed in November 2010.
- Gross, R., A. Acquisti, and H. Heinz. 2005. Information revelation and privacy in online social networks. WPES. Alexandria, VA, U.S.A.
- Hildebrandt, M. 2006. Profiling: from data to knowledge. *DuD: Datenschutz und Datensicherheit* 30(9).
- Johnson, C. 2009. Project Gaydar. http://www.boston.com/bostonglobe/ideas/articles/2009/09/20/project_gaydar_an_mit_experiment_raises_new_questions_about_online_privacy/. Accessed in November 2010.
- Kamkar, S. 2010. Evercookie—never forget. <http://samy.pl/evercookie/>. Accessed in November 2010.
- Kirkpatrick, M. 2010. Google CEO Schmidt: “people aren’t ready for the technology revolution”,. http://www.readwriteweb.com/archives/google_ceo_schmidt_people_arent_ready_for_the_tech.php. Accessed in November 2010.
- Krishnamurthy, B., and C. Wills 2008. Characterizing privacy in online social networks. In WOSN’08: Proceedings of the first workshop on Online social networks. Seattle, WA, U.S.A.

- Krishnamurthy, B., and C. Wills 2009a. On the leakage of personally identifiable information via online social networks. In WOSN' 09: the second workshop on Online social networks. Barcelona, Spain.
- Krishnamurthy, B., and C. Wills. 2009b. Privacy diffusion on the web: a longitudinal perspective. In WWW' 09: Proceedings of the 18th international conference on World wide web. ACM. Madrid, Spain.
- Krishnamurthy, B., and C. Wills. 2009c. Privacy diffusion on the web: A longitudinal perspective (updated graphs). <http://www.ftc.gov/os/comments/privacyroundtable/544506-00009.pdf>. Accessed in November 2010.
- Krishnamurthy, B., and C. Wills. 2010. Privacy leakage in mobile online social networks. In WOSN' 10: Proceedings of the third workshop on Online social networks. Boston, MA, U.S.A.
- Krumm, J. 2010. Ubiquitous advertising: The killer application for the 21st century. IEEE Pervasive Computing.
- Le Métayer, D. 2010. Privacy by design: A matter of choice. In *Data protection in a profiled world*, ed. S. Gutwirth, Y. Poullet, P. De Hert, 323. Verlag: Springer.
- Macmanus, M. 2009. A guide to recommender systems. http://www.readwriteweb.com/archives/recommender_systems.php. Accessed in November 2010.
- McKinley, K. 2008. *Cleaning up after cookies*. Technical report, iSEC PARTNERS. https://www.isecpartners.com/files/iSEC_Cleaning_Up_After_Cookies.pdf. Accessed in November 2010.
- Miluzzo, E., N. Lane, K. Fodor, R. Peterson, H. Lu, M. Musolesi, S. B. Eis, X. Zheng, S. EisenMan, and A. Campbell 2008. Sensing meets mobile social networks: The design, implementation and evaluation of the cenceme application. Proceedings 6th ACM Conference on Embedded Networked Sensor Systems (SenSys' 08). Raleigh, NC, U.S.A.
- Narayanan, A. 2010. Do not track explained. <http://33bits.org/2010/09/20/do-not-track-explained/>. Accessed in November 2010.
- Raphael, J. R. 2011. Apple vs. Android location tracking: Time for some truth. http://blogs.computerworld.com/18190/apple_android_location_tracking. Accessed in July 2011.
- Schneier, B. 2009. Architecture of privacy. *IEEE Security and Privacy*.
- Schoen, S. 2009. New cookie technologies: Harder to see and remove, widely used to track you. <http://www.eff.org/deeplinks/2009/09/new-cookie-technologies-harder-see-and-remove-wide>. Accessed in November 2010.
- Zheleva, E., and L. Getoor. 2009. To join or not to join: The illusion of privacy in social networks with mixed public and private user profiles. In International World Wide Web Conference (WWW). Madrid, Spain.
- Zhong, G., I. Goldberg, and U. Hengartner. 2007. Louis, lester and pierre: Three protocols for location privacy. Proceedings of the 2007 Privacy Enhancing Technologies Symposium (PETS). Ottawa, Canada.

European Data Protection: In Good Health?

Gutwirth, S.; Leenes, R.; de Hert, P.; Pouillet, Y. (Eds.)

2012, XVIII, 363 p. 20 illus., 10 illus. in color., Hardcover

ISBN: 978-94-007-2902-5