International Working Group
on Data Protection
in Telecommunications

675.46.13

Working Paper on

Web Tracking and Privacy:

Respect for context, transparency and control remains essential

53rd meeting, 15-16 April 2013, Prague (Czech Republic)

Introduction

1. This paper is based on a foundation of respect for the fundamental rights and freedoms of Web users. Although it does not focus on specific technical remedies the paper does assume that the technical action of Web tracking must be lawful, appropriate and that it must operate within a strict framework of those rights. The principles of choice and control - claimed by much of industry - sit at the core of this framework, and those principles must be enacted with precision upon the pillars of clarity, transparency and accountability. The justification for the imposition of Web tracking is not self evident and thus industry and other tracking exponents must continually strive to explore solutions that bring this activity not just squarely within the framework of fundamental rights and privacy, but also in line with the imperative of Privacy by Design.

2. In this working paper, the Working Group addresses the issue of Web Tracking and Privacy. Although no clear definition exists, we will refer to a definition of Web Tracking[1] as the collection, analysis and application of data on user activity from a computer or device while using various services of the Information Society (hereinafter: the Web)[2] in order to combine and analyze it for different purposes, from charitable and philanthropic to commercial. We consider various forms of market research to fall within this definition of Web Tracking, for example outreach measurement (the degree to which users are served with ads across the Web), engagement measurement (the degree to which users interact with services across the Web) and audience measurement (the degree to which micro profiles can be derived from users interacting with services across the Web).[3]

[1] Cf. van Eijk (2012), The DNA of OBA: unique identifiers, URL: http://www.campusdenhaag.nl/crk/publicaties/robvaneijk.html#definition-of-web-tracking.

[2] Note that with IP-based technology becoming the backbone of the information society, and integrating many other former "stand alone" technologies ("Convergence"), this may well encompass the use of a telephone (IP telephony), television (IPTV), reading digital newspapers, or any other media consumption using digital technologies (including reading an e-book). For a detailed discussion of the resulting privacy risks cf. theWorking Paper on Privacy Issues in the Distribution of Digital Media Content and Digital Television (Berlin, 4./5.09.2007) of this Group; URL: http://www.datenschutz-berlin.de/attachments/349/digit_en.pdf

[3] JICWEBS Reporting Standards, URL: http://www.abc.org.uk/PageFiles/50/Web Traffic Audit Rules and Guidance Notes version2 March 2013 master.pdf

Secretariat
Berliner Beauftragter für
Datenschutz und Informationsfreiheit
An der Urania 4- 10
D-10787 Berlin
Phone +49 / 30 / 13889 0
Fax: +49 / 30 / 215 5050

E-Mail:
IWGDPT@datenschutz-berlin.de

Internet:
http://www.berlin-privacy-group.org

Scope of the Working Paper

3. The paper is addressed to all providers of web sites as well as software developers and service providers offering or using tracking technology. This paper discusses the development of tracking technologies and their possible impact on the privacy of citizens. This paper deals with digital traces left behind when using various services of the Information Society with a Web Browser, including unique identifiers derived from non-cookie based techniques.[4] This includes Web Browsers on other devices, for example smart mobile devices and smart televisions.

4. This paper does not deal with specific additional risks which may stem from the advent of apps on mobile devices.[5] Nevertheless the principles in this paper should also be applied for tracking mechanisms used in other services.

5. This paper is not about how protective measures can be implemented (e.g., legal requirements for consent). Note that while in some jurisdictions, depending on the purpose of Web Tracking, explicit consent (opt-in) is required, in other jurisdictions, an opt-out for Web Tracking will be considered valid to satisfy the legal framework if certain conditions are met. These include, among other things: adequate notification of processing; transparency in the notification; notification at or before the time of collection; and simple, effective and persistent opt-out methods. A number of restrictions may also be in place, including limiting the processing sensitive information such as information on health, information on political or philosophical beliefs and the prevention of the tracking of children.

Background

6. The technical possibilities of monitoring the activities of users on websites have multiplied over the past decade and the emerging "Information Society" has seen several sea changes since then.[6] Web tracking developed from very modest beginnings - when single providers of online services started to monitor their users to find out whether a particular user had been there before and what this user had been doing - into an almost panoptical vision of marketers more recently. In this vision, the marketer seems to be able to monitor every single aspect of the behaviour of an identifiable user across websites. This could potentially become a complete history of the entire Internet usage of a data subject (literally from the cradle to the grave), and could be enriched with profile data from the former "offline world" (including any aspect of our lives data brokers have information about, including financial information as well as information on, for example, leisure, health, political and/ or religious opinions, location information).[7]

7. This development - while greeted and fostered by marketers and other interested parties from the broader business community, and assisted by some policymakers at the national and regional levels - holds an unprecedented risk for the privacy of all citizens in an information society. The worst case scenario is that it would turn the world as we know it into a global panopticon. The offline equivalent would be to have somebody unknown to us constantly looking over our shoulders no matter where we are (in the streets or in the seeming privacy of our homes), or what we do (watching TV, shopping online, reading

[4] For example, passive fingerprinting techniques based on hashing the HTTP user agent and/or the IP address of the originating browser.

[5] See, for example, Opinion 02/2013 on apps on smart devices WP 202 issued by the Article 29 Working Party (Art. 29 WP), URL: http://ec.europa.eu/justice/data-protection/article-29/documentation/opinionrecommendation/files/2013/wp202_en.pdf

[6] The literature review on Web privacy measurement, which has been produced as an outcome of the Conference on Web Privacy Measurement (WPM) gives a more elaborate view on the technologies used for tracking, URL: http://www.law.berkeley.edu/12633.htm

[7] In Customer Relationship Management (CRM) the common terms are Customer Lifetime and Customer Lifetime Value.

newspapers, and even more intimate activities), and without knowing when he is looking, and when he isn't.[8]

8. The possible repercussions of such a development are evident and not to be underestimated with respect to their potential gravity. It may annul and do away with some of the core principles of privacy - and notably transparency and control for the individual.[9] To put it more bluntly, this might be the end of the (privacy) world as we know it.

9. The promoters of this vision, on the other hand, claim that these risks either do not exist at all, or that they have tried to address and mitigate these risks at least in part. There is strong resistance from some stakeholders from industry against recognizing that unique identifiers in Web data are personal information. One claim often put forward is that much of the data in use has been de-identified (i.e., anonymised), and that once this has been done, the data is no longer about a person and would therefore not pose a risk to the privacy of citizens. It is also claimed that any behavioural data are linked to machines only and can - this is the claim - in very many instances not be traced back to an individual at all.

10. However, these claims have no scientific proof whatsoever, and ignore the fact that machines - and especially smart phones - are becoming more and more personal devices and allow for an easy link to any given individual user. Traces can also increasingly be linked across different devices. There is also scientific proof that many seemingly anonymous data (e.g., location information of cell phones) can be traced back (i.e., be de-anonymised) to any given user if the database and the timeframe are sufficiently broad. Even worse, more recent academic work suggests that it is impossible in principle to keep "anonymous" data from being de-anonymised if the time slice depicting any given behaviour is sufficiently big (i.e., it is conceptually impossible to guarantee that "anonymous" data cannot be traced back to an individual over time). If this holds true, it is a game changer and will make a couple of core assumptions about how uses of different types of data may or may not affect the privacy of individuals useless.[10]

11. In addition, and on a slightly different note, practical daily knowledge also adds to questioning the claims made by industry. While ads may well be addressed to a machine at the technical level, it is not the machine which in the end buys the proverbial beautiful pair of red shoes - it is an individual. Thus, the claim that the processing of behavioural data for marketing is directed "only" at machines in the first place may well be seen as an attempt to blur our vision as societies on the gravity of the problem, when in reality the individual and not the machine is the only instance that can make all such tracking operations a "success" for its proponents (i.e., when the red shoes are finally being bought).

A short history of monitoring technologies

12. In trying to trace back the development described above to its modest beginnings, one milestone we find is the development of "cookie technology" almost 20 years ago. HTTP Cookies were introduced in 1994, first and foremost to solve the "small" problem of reliably implementing a virtual shopping cart. Due to the mostly stateless nature of the Hypertext Transfer Protocol (HTTP), user agents were not able to retain state information until then. Retaining state information was crucial for the virtual shopping cart in order to remember selected items during the shopping experience. Transparency was already then a privacy issue, because the use of cookies was not conveyed to the ordinary user. At the time, cookies

[8] To make things even worse, this modernist version of the panopticon would record every single move of any given individual at any given moment in time no matter whether the guard is watching or not.
9

[9] Tracking as a technology is not transparent. At the technical level, in many cases, the pixels (e.g., web beacons) and mini webpages (e.g., iFrames) are invisible to the human eye

[10] Cf. Ohm, Paul: Broken Promises of Privacy: Responding to the Surprising Failure of Anonymization, August 2009. URL: http://papers.ssrn.com/sol3/papers.cfm?abstract_id=1450006

were enabled by default in the browser settings and the user was not notified about the use of cookies.[11]

13. To mitigate the privacy and security risk of leaking cookie information to other sites the same origin policy was implemented. This policy meant that cookies could only be read by the same domain that set them. However, it is important to note that recommendations through the World Wide Web Consortium (W3C) propose a new standard, Cross Origin Resource Sharing (CORS)[12] which will permit the sharing of information across specified domains. Although CORS is a voluntary standard, it conflicts with the same origin policy.

14. In 1998, this group[13] addressed various privacy issues connected to the systematic collection or use of personal data on the Web.[14] In its working paper, it addressed P3P (Platform for Privacy Preferences Project), a protocol developed by W3C, which was designed to block third party cookies unless the website the user visited offered a user acceptable P3P policy.[15] However, only one major browser manufacturer implemented the standard. As a result, P3P has not been adopted widely on the Web.

15. Third party cookies have become the lifeblood of the complex digital ad industry. In 2008 marketing executives of Web Tracking companies discussed the future of analytics and site statistics. The future, five years ahead, was envisioned to be an integration of traditional sitevisit statistics (hereinafter: First and Third Party Analytics) and analytics data from other services on the Web including, for example, video, widgets, social networking, gaming and search engines (hereinafter: Web Analytics).[16]

16. Today, Web Analytics Data represents a new form of economic value. While this group does not question the benefits that measuring consumer behaviour may bring for (real-time) online behavioural advertising (OBA), it firmly believes that such practice must not be carried out at the expense of individuals' rights to privacy and data protection.

Web Tracking

17. Web Tracking involves the collection and subsequent retention, use or sharing of data on individual online behaviour across multiple websites by the use of cookies, JavaScript or any kind of device fingerprinting. Web Tracking technology enables a constant flow of real-time information about users, such as registration data, search activities, behavioural data, site visit statistics and conversion data reflecting how a user responded to individual offers. These data can be used to infer users' interests, political opinions or medical conditions. These data can be processed with the purpose to evaluate, treat in a certain way or influence the status or behaviour of an individual. Data about individual behaviour drives business decisions based on customer profiles. Buying intent may be derived from a person's

---

[11] RFC 2109, HTTP State Management Mechanism, URL: https://tools.ietf.org/html/rfc2109. Note that current flavors of cookie storage technology include for example flash cookies and the LSOs (Local Shared Objects) used in HTML5 with matching values.

[12] Cross-Origin Resource Sharing, URL: http://www.w3.org/TR/cors/ ; W3C "Candidate recommendation" status since 29 January 2013 (viewed on 30 May 2013).

[13] International Working Group on Data Protection in Telecommunications

[14] Common Position on Essentials for privacy-enhancing technologies (e.g. P3P) on the World Wide Web (Hong Kong, 15.04.1998), URL: http://www.datenschutz-berlin.de/attachments/178/priv_en.pdf .

[15] The Platform for Privacy Preferences Project (P3P) enables Websites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents will allow users to be informed of site practices (in both machine- and human-readable formats) and to automate decision-making based on these practices when appropriate. Thus users need not read the privacy policies at every site they visit. , URL: http://www.w3.org/P3P/ .

[16] Omma Global Measurement 3.0, URL: http://www.webmetricsguru.com/archives/2008/09/measurement-30-on-the-next-5-years-omma-global-day-2/.

presumed digital identity. The value of a potential customer is related to the chance to convince him to buy a product.

18. Web Tracking technology is present on mobile devices. A smart mobile device is unlikely to be shared between individuals, therefore making the link between the device and the individual stronger than with, for instance, desktop computers. Mobile devices contain unique device identifiers such as advertising specific identifiers,[17] the Unique Device ID (UDID), Media Access Control (MAC) address, Bluetooth MAC address, Near Field Communications (NFC) MAC address, International Mobile Subscriber Identifier (IMSI, a unique SIM card number) and the International Mobile Equipment Identifier (IMEI). These identifiers cannot be changed by ordinary users. In addition to unique identifiers, smart mobile devices may contain a rich set of data such as user name, password, age, gender, and address book. Smart mobile devices can expose accurate behavioural data on the whereabouts of a user. Precise geolocation data is readily accessible for browsers on smart mobile devices.

19. Web Tracking technology is deployed in various ways. A digital data trail may result from unintentional or unwilling disclosure of data, and may result in unnecessary disclosure of (personal) data. There are multiple ways to generate a digital data trail. For example, a campaign manager for digital ads could assign a unique identifier to the user, browser or device. Another way is to personalize referral information by adding audience segment information (micro profiles) while surfing the Web, so other sites participating in the campaign can track the user, browser or device too. A third example is by correlating unique identifiers with data collected from past visits on a specific site. A fourth example is that Web Tracking for a campaign can also take place by combining new tracking data (about a user, browser or device data) with data previously collected on a specific site, or data obtained from another (third) party. A final example involves the use of cookie matching services that connect digital trails from the same user, browser or device with the use of different parts of the Web.[18]

20. Web Tracking consists of several automated steps, starting with the collection of Web data, the retention of these data, and the use of the data. By recombination, correlation and decontextualization, Web data can be used to construct very detailed predictive profiles of individual behaviour. Finally, Web Tracking leads to the actual application of the profile to an individual.[19]

21. Data can be stored in graph databases by various services on the Web.[20] The graph structure enables the emergence of behavioural patterns that would otherwise remain undetected. Web Tracking data in a graph can create meaningful patterns about user behaviour by itself or when combined with other data from various sources. For example, while individual unique identifiers connected directly or indirectly to a user or computer may expose little information about the casual surfer, the collection of unique identifiers reveals a pervasive view of someone's habits and browsing behaviour on the Internet. The collection of unique identifiers can be used to construct a digital identity.

Web tracking and the right to privacy and data protection of the individual

[17] For example, to be able to perform frequency capping (control of the number of times a user has seen an ad), to deliver behavioral ads, and to measure the reach and effectiveness of an advertising campaign.
[18] See for example URL: https://developers.google.com/ad-exchange/rtb/cookie-guide#what-is .
[19] Cf. also Recommendation CM/Rec(2010)13 of the Council of Europe on the protection of individuals with regard to automatic processing of personal data in the context of profiling.
[20] A graph is based on graph theory which is a mathematical approach to model pairwise relations between objects. A graph database stores graphs which are essentially structures with nodes, edges, and properties. Theproperties may contain meta information about the nodes and edges:

22. A key principle across a broad range of international legislative frameworks is the right to privacy that the Internet user has regardless of technology. Key elements are transparency, control and respect for context. The fact that users are unaware that they are being tracked is a privacy risk. Web Tracking as a process utilises a number of technical tools which limit the opportunity for users to be notified. For example, pixels (e.g., web beacons) and mini web pages (e.g., iFrames) are invisible to the human eye and inclusion in a web page will initiate an automatic HTTP request including the opportunity to set and access cookies containing unique identifiers.

23. Many web tracking technologies have been developed and deployed in business without providing information to the users whose data is being collected and without giving them any choice. User signals that could be understood as expressing objection to tracking have been disregarded and technical mechanisms against some tracking mechanisms have been actively circumvented, for example, by re-spawning deleted cookies, (passive) fingerprinting, and circumventing browser settings. Only when these behaviours were detected and were publicly criticized did the interested parties accept their obligation to respect users' free will. In such cases, sometimes opt-out schemes have been added after the fact, often leading to clumsy mechanisms of limited usefulness for the user. These cases have caused great damage to the users' trust in the reliability and honesty of all web service providers and undermine the healthy development of innovative web services.

24. Web Tracking constitutes processing of personal data in many jurisdictions due to the fact that the technology enables the individualization or identification[21] of users and/or making automated decisions about them. An example of such practice might be automatic decisions engines with algorithms in real time bidding platforms for personalized behavioural advertising.

25. There is strong resistance from some interested stakeholders against classifying unique identifiers in Web data as personal information. One claim often put forward is that once data has been de-identified[22], the data is no longer about a person. It should, however, be clear that a "purpose" element can also be responsible for the fact that information "relates" to a certain person or is about a person.[23]

The potential impact (or lack of impact) of "Do Not Track" (DNT) - a case study

26. In September 2011, the W3C chartered the Tracking Protection Working Group[24]. The group is working on a Do Not Track (DNT) standard. All major browsers have committed themselves to implement the standard (and most have already so implemented the HTTP header), however there remains, amongst those stakeholders who will honour the DNT:1 request[25], an open discussion on parts of the voluntary standard. Some stakeholders have indicated they will not honour the DNT flag for various reasons. The overall success of DNT is tied to

---

[21] Recital 26 of the general Data Protection Directive 95/46/EC: Whereas the principles of protection must apply to any information concerning an identified or identifiable person; whereas, to determine whether a person is identifiable, account should be taken of all the means likely reasonably to be used either by the controller or by any other person to identify the said person (...), URL: http://eurlex.europa.eu/LexUriServ/LexUriServ.do?uri=CELEX:31995L0046:en:HTML.

[22] De-identification means deleting, modifing, aggregating, anonymizing or otherwise manipulating data.

[23] Opinion 4/2007 on the concept of personal data (WP136), p. 10 URL: http://ec.europa.eu/justice/policies/privacy/workinggroup/wpdocs/2007_en.htm .

[24] The mission of the Tracking Protection Working Group is to improve user privacy and user control by defining mechanisms for expressing user preferences around Web tracking and for blocking or allowing Web tracking elements, URL: http://www.w3.org/2011/tracking-protection/charter.

[25] In the current draft DNT standard, sending "0" signals that tracking is fine, while "1" indicates a wish NOT to be tracked.

the actual honouring of the DNT flag by receiving organizations and the factual adoption of the DNT standard throughout the Web by all stakeholders.

27. The default settings of DNT and the default actions by the Web Tracking organisation remain crucial once again. For DNT to be an effective instrument to provide user control, it is crucial that those performing Web Tracking can be certain that the DNT signal which they receive is a true indication of the user's wishes. In the absence of fully informed user choice, a Web Tracking organisation must assume that a user is not aware of Web Tracking and therefore assume the default position as if they had received a DNT:1 signal, which indicates a wish from the user not wanting to be tracked.

28. Any technology used for Web Tracking purposes must be proportionate. Data protection principles used worldwide are based on the notion that data should be collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes. Data processing should be adequate, relevant and not excessive in relation to the purposes for which they are collected and/or further processed.

29. Finally, any technology must be "court-proof" if it is to contribute to serving the protection of privacy. DNT is in danger of remaining a tool through which a user may express wishes to service providers in the information society, without being an effective granular dialogue instrument. This leaves the user himself or any public (or private) body being chartered with enforcing those wishes or rules (and including corresponding legal obligations to honour any such choices made by an individual) empty handed vis-à-vis those providers. Some industry stakeholders try to defend the position that DNT does not constitute an obligation to respect such a wish. While this interpretation is more than doubtful, the fact remains that it is difficult to prove whether such a wish has been respected and or been disregarded.[26] In other words, from an enforcement perspective, DNT could remain a sugar pill instead of being a proper cure and would as such be useless.

Recommendations

30. Unchecked Web Tracking may change the balance between service providers and individuals, including with respect to privacy protection. The Working group underlines that context, transparency and control remain crucial elements in the context of Web Tracking.

31. In order to contribute to addressing the risks for the privacy of the individual, the Working Group makes the following recommendations to the different stakeholders who have a part to play in the Web Tracking ecosystem.

Re-introduce respect for context and purpose limitation as core principles for any use of personal data:

- incorporate precautionary principles in any (automated) data collection, processing and sharing practices, so that data collected in one context cannot be applied in another context; and

- inform about the purpose of data collection in advance and do not change the purpose without renewed information and choice.

Bring back transparency:

[26] External audit might play an important role in addressing at least parts of the problems described above, but would on the other hand add even further complexity to the ecosystem.

- Refrain from the use of invisible tracking elements;
- As a minimum, notify the user in an intelligible way when the user agent is about to send/receive a Web Tracking identifier to/from the origin/destination server;
- Display an indicator noticeable enough to the user[27] whenever Web Tracking is in progress; and
- Make an indication that Web Tracking is in progress also available to special groups of users, including the visually impaired.

Put the user back into control:

- implement mechanisms that allow users to exercise their right to privacy and data protection on the Web and do not deploy any (new) tracking mechanisms that do not have a user control mechanism; offer users an explicit choice regarding tracking - when browser software is to be installed, activated or updated, there must be a user choice;
- if the browser does not provide a user interface, the default setting should be such that the user is not tracked;
- give users the opportunity to reconsider their choice and change settings after the initial decision and at any time; let the user examine the (automated) choices that have been made with regards to Web Tracking in an easy way; and remind the user that choices regarding the (automated) settings for Web Tracking can be revoked at any time and make sure that a revision of any such choices is technically possible in an easy way that does not put any undue burden on the individual;
- honour requests when the user agent is signalling that it does not want to be tracked;
- refrain from (passive) fingerprinting, for example by mining user generated data (such as service configurations, or user agent strings) in order to derive a unique user identifier (device fingerprint) when a user has expressed not wanting to be tracked; and
- ensure that the application of any technology devised to let users make choices is auditable and can be enforced by the competent private or public bodies chartered with enforcing rules, and especially those enshrined in the different existing legal frameworks which provide the foundation of the protection of privacy of the individual in many jurisdictions across the globe.

[27] Special consideration must be given to ensure that no group of web users are treated less favourably or are otherwise discriminated against, for example, as a result of a disability.