Privacy, cookies, and web analytics

Dr. Lionel Khalil, Director of the Office of Institutional Research and Assessment, Notre Dame University, Louaize, Lebanon

Abstract: *The generalization of monitoring data records has established the need to ensure safety and protection of privacy rights. There are two different trends in managing cookies for web analytics: egocentric and sociocentric approaches, which, when used simultaneously, are considered too intrusive and are limited by most regulators. In fact, it is difficult to achieve true web anonymity because there is a tradeoff between functionality and privacy. It is not easy to give informed consent clarifying when and what to disclose, what is useful, why it should be disclosed, and with whom it should be shared (relatives, coworkers, friends). In practice, private data are vulnerable. Even though most applications that access users' address books have good intentions, several applications are solely designed to steal phonebook contacts. Considering that the practice of uploading address book data is so widespread, the legislative answers to those breach of privacy are not efficient , and the uncertainty is enough to make even the most trusting of people paranoid. There are several legislative actions to be taken to mitigate this growing threat to privacy and to inform customers on the main dangers of those applications.*

The generalization of monitoring data records has established the need to ensure safety and protection of privacy rights. Claims of privacy violation increase with public discomfort over the lack of appropriateness between the level of data recorded and the incidental usage of this data. It is difficult to define an acceptable general level of privacy violation for all because this level is specific to each culture.

The common definition of personal information according to OECD Privacy Principles[1] is any data that can lead to the identification of an individual, either alone or when combined with other identifying information. In the case of web analytics, information collected may be considered within this definition of personal information. Here is the debate.

The use of web analytics has exploded with the development of social media. Web analytics helps software companies or website owners to better know their customers, focus their commercial efforts, and build better relationships with customers. In practice, web analytics is based on collecting personal data from the customer with or without informed consent. The information collected depends on the application and the device (mobile or PC) but falls in the following categories:

- Personal data such as name, device ID number, IP address, or location
- Data stored in the device (PC or mobile) such as contacts or photos
- Potential competitor information that can be found in the history of navigation and competitors' cookies
- Behavioral data such as the user's pattern of clicking on a page or keywords used in search engines

Thanks to web analytics, website owners can know their customers' personal data, the contacts of their friends, if they have visited competitors, and their way of using the website. That information helps owners propose better services and place appropriately customized advertising on the screen. For example, Gmail reads the emails received in their customers' mailboxes and proposes advertising based on the mail content. Most of the applications on mobile use web analytics without the express consent of the user.

[1] OECD Privacy Principles, http://oecdprivacy.org/

There are two different trends in web analytics[2]. The first one, an egocentric approach, reflects the traditional forensic trend of identifying the profiles and the behaviors of individuals, trailing them, their friends, their connections, and their daily activities to create a sound portrait of them, and using them as viral vectors to promote a product or a service. The second trend, a sociocentric approach, is a normative trend to characterize the imprint of a person by his or her behavior and reduce this personal information's imprint into predefined, normalized, and schematic categorical sets that might be used to target the individual specifically with a generic product fitting his or her current need.

In Part I, we will discuss the conflict between profiling people and their right for anonymity. In Part II, we will present the need for a legal framework to monitor and limit the use of web trackers such as cookies.

Part I – Profiling a person and the right for anonymity

Users tend to have very contradictory behavior when they face the issue of privacy. On one hand, many users publicize each and every part of their private life, but on the other hand, they have a general sense of privacy violation.

1 – The right of anonymity

In general, does the law enforce anonymity? Even the right to freedom of movement, expressed by Article 13 of the Universal Declaration of Human Rights and recognized in many countries, does not necessarily guaranty the freedom of movement to be exercised under anonymity. There are a few legal statuses in several countries that allow full anonymity. In the health sector, there is donor anonymity of body organs, elective abortion, and anonymous child birth. In the press, the reporter's privilege has always been legally defined as the professional right to maintain the anonymity of sources because this relationship furthers the free flow of information to the public. In special cases, some auxiliaries of justice and some witnesses can benefit from anonymity as well. In literature, an anonymous work is a work published by an author who decides to remain anonymous for a variety of reasons and publishes the work either under a pseudonym or under no name at all. This collection of specific examples is not exhaustive, but it shows that anonymity is seen as an exception more than a general principle.

There is also a significant difference between the lawful collection of data under confidentiality and real, global anonymity. Entities such as medical professionals, employers, state bodies, lawyers, and accountants have the authority to collect data on people, and such data are indeed kept secret. This confidentiality is the consequence of the principle of finality for any collection of personal data. The responsibility of confidentiality is bound by professional obligation. Rather than recognizing a full right for anonymity, the general principles of law recognize a general right for privacy in any interaction with other bodies of the society. This right for privacy seems a fair balance between the need to identify citizens in their interaction with society and the legitimate need for anonymity of each person.

[2] Facebook, Twitter et les autres... p174, by Christine Balagué, David Fayon, PEARSON VILLAGE MONDIAL (26 février 2010) ISBN-13: 978-2744064197

Privacy is widely protected since 1980 by states' laws and more generally thanks to the guidance of the model law from the OECD Privacy Guidelines of 1980[3] and the APEC Privacy Framework of 2004.

In May 2014, a major change in internet privacy came from a ruling of the European Union's highest court, the European Court of Justice, which ruled that internet search companies must respect a "right to be forgotten." This "right to be forgotten" requires anonymity in data that have been collected in the past, but this right does NOT imply that the collection of the data was illegal. The main issue affected by this ruling concerns press reports of names of convicted individuals who have rightfully served their time. Several years after their release, those convicted persons can request that their name be removed from the online archives of newspapers. There is balance between a legitimate need to save information and the resulting prejudice when dissemination of this information becomes harmful for the formerly convicted individuals trying to begin a peaceful life.

2 – The use and abuse of personal information to profile individuals

Social network businesses are centered on the exploitation of personal information. A specific example of this is Truecaller, which commercializes the aggregation of personal contacts from users' address books. Phone owners do require their contacts' permission to put their name and numbers in their address book contacts. However, by aggregating the personal contacts of their users, Truecaller has legitimate access to a wide phone database. Any person can request to be removed from the shared phone database of Truecaller, but by default, any name and number available from TrueCaller users are in the shared database.

Another example of commercial exploitation of personal information is the 2012 Path Scandal[4], which involved the right to access address books of clients. This right is essential to most "Find my friends on this service" features. The popular Path app was caught uploading and permanently storing people's entire address books on Path's servers without requesting any permission.

In a recent study from Peter Gilbert et al.[5], Duke University has shown Apps' general violation of privacy by employing a traffic-monitoring utility called mitmproxy to observe the data flowing between apps and the internet. Out of 30 popular apps, 15 send information such as personal contacts, microphone capture, camera capture, or location without the consent of the user to the owner of the application.

App developers and owners defend themselves against claims of privacy violation with the following line of defense: First, app developers and owners protect themselves with a data privacy policy or an end-user license agreement resulting in explicit but mostly uninformed consent. Second, they argue that the access to the personal information is officially and technically allowed by Apple or Android. Third, they argue that web analytics robots access and read the personal information, but robots do not copy the personal information, and there is no human access to the personal information.

---

[3] OECD guidelines on the protection of privacy and transborder flows of personal data, by OECD (1980); APEC Privacy Framework, APEC (2004);

[4] *Path CEO apologizes for address book uploading, deletes all user data, and updates app with privacy controls*, By Nathan Ingraham on February 8, 2012,http://www.theverge.com/2012/2/8/2785217/path-ios-address-book-upload-ceo-apology

[5] *Automating Privacy Testing of Smartphone Applications*, by Peter Gilbert, Byung-Gon Chun, Landon P. Cox, Jaeyeon Jung Duke University, Intel Labs retrieved from http://www.cs.duke.edu/~lpcox/TR-CS-2011-02.pdf

Here is the debate about permission and consent. If the door of your garden is not locked, does it imply that someone can rightfully enter in your garden to take photos and come into your house?

For Daniel J. Solove[6], privacy self-management is the right to notice, access, and consent regarding the collection, use, and disclosure of personal data. Even well-informed individuals face structural problems that reduce their capacity to appropriately manage their privacy. Often an individual is limited to managing separate privacy rights for each entity (website, apps, etc.) that collects and uses personal data. Moreover, an individual person is often not able to conceptualize the real threat to privacy through the aggregation of several pieces of information over a period of time by different entities. Finally, it is difficult to assess harm in the long term while balancing it with immediate benefits.

In contrast, Hoofnagle et al[7] describes the interventions to support consumer privacy interests as paternalistic judgments that individuals cannot make proper decisions for themselves. Merely giving consumers some legal or technical mechanism to block such data collection and tracking is paternalistic because it intervenes in the natural market ecosystem between users and websites.

There is a tradeoff between functionality and privacy. Nevertheless, the complex dilemma with consent remains between the idea that express consent to collect, use, and disclose personal data is often not very meaningful and the idea that paternalistic measures deny people the freedom to make consensual choices about their data.

Part II – The need for a legal framework to monitor and limit the use of web trackers such as cookies

Cookies were designed around 1990 and were initially designed to keep track of browser-server interaction during a session on a website or between two sessions. In 1995, the Internet Engineering Task Force (IETF) initiated a standardization process for cookies[8]. In 2000, the IETF published the RFC 29653 "HTTP State Management Mechanism"[9], which specified a way to create a session with HTTP requests and responses.

This began to raise both security and privacy concerns when third parties started to read them. There is limited support for confidentiality, integrity, and authentication in the way cookies are used, and because they are used to store core information about interaction with the website, the possibilities for misusing cookies are very real and are sometimes exploited. They can be misused mainly for profiling.

---

[6] John Marshall Harlan Research Professor of Law, George Washington University Law School, Solove, Daniel J., Privacy Self-Management and the Consent Dilemma (November 4, 2012). 126 Harvard Law Review 1880 (2013); GWU Legal Studies Research Paper No. 2012-141; GWU Law School Public Law Research Paper No. 2012-141. Available at SSRN: http://ssrn.com/abstract=2171018

[7] Behavioral Advertising: The Offer You Cannot Refuse (August 28, 2012), by Hoofnagle, Chris Jay and Soltani, Ashkan and Good, Nathan and Wambach, Dietrich James and Ayenson, Mika. Harvard Law & Policy Review 273 (2012); UC Berkeley Public Law Research Paper No. 2137601. Available at SSRN: http://ssrn.com/abstract=2137601

[8] *HTTP Cookies: Standards, privacy, and politics,* by David Kristol, ACM Transactions on Internet Technology, 1(2), 2001, pp.151–198, available at: http://www.cs.stevens.edu/~nicolosi/classes/sp10-cspriv/ref5-1.pdf

[9] HTTP State Management Mechanism, IETF RFC 2965, published in 2000, available at: http://www.ietf.org/rfc/rfc2965.txt

There are alternative technologies to cookies[10] that are more difficult for users to detect and block. The first technique stores a unique identifier on the user's computer that advertisers can use to track individuals (such as ETags, Flash cookies, HTML5 local storage, and Evercookies). The other technique identifies users by their unique imprints, which are based on the serial numbers of the browsers used and other characteristics such as fonts and pictures.

In October 2009, the European legal framework amended the European Union's e-Privacy Directive (2002/58/EC) to require website users to opt-in to tracking cookies. Article 5(3) states: "the storing of information, or the gaining of access to information already stored, in the terminal equipment of a subscriber or user is only allowed on condition that the subscriber or user concerned has given his or her consent, having been provided with clear and comprehensive information." Unfortunately, few E.U. countries have implemented the amended e-Privacy Directive. France has amended in August 2011 the Article II-32 of the Act of 6 January 1978, to transpose the e-Privacy Directive[11]. Under the UK 98'Data Protection Act, tracers (cookies or other) require obtaining consent and may not be posted or read on a user's device as long as the person has not consented.

In practice, the limitations proposed by the French CNIL severely limit the use and abuse of tracking. The data collected cannot be cross-checked with other sources such as client files. The cookies deposited must be used solely for the production of anonymous statistics. Therefore, even if the user consents, the data collected cannot be reused in another application.

These constraints do not allow a website to mix egocentric and sociocentric usages of cookies. Let us look at several practical applications of this CNIL limitation. In the first scenario, a customer reads an FAQ page several times to understand a feature. There is no right to proactively advise him or offer personalized assistance. In the second scenario, a user has a bug visiting a website, but the website is not allowed to use cookies to know what browser was used and in which conditions it was used. In the third scenario, a website owner cannot crosscheck the usage of his website with the type of user subscription to propose a more adapted service to the user.

This limitation imposed on cookie usage by the CNIL against using the collected data in both egocentric and sociocentric ways limits the capacity to improve the products and services proposed by the website.

It is difficult to achieve true anonymity. There is a tradeoff between functionality and privacy. It is not easy to give informed consent clarifying why, when, and what to disclose. People want to disclose only what is useful and decide with whom they want to disclose it (relatives, coworkers, friends). Nevertheless, the real privacy harm is the aggregation of private data in the long term. Most of the applications that access users' address books have good intentions, but several applications are solely designed to steal contacts. The practice of apps uploading address book data is currently very common, and because the real usage of the address data book is unknown, users can develop paranoid attitudes. There are several actions to be taken in observing this growing threat to privacy and in informing customers on the main dangers of those applications.

---

[10] Behavioral Advertising: The Offer You Cannot Refuse (August 28, 2012), by Hoofnagle, Chris Jay and Soltani, Ashkan and Good, Nathan and Wambach, Dietrich James and Ayenson, Mika. Harvard Law & Policy Review 273 (2012); UC Berkeley Public Law Research Paper No. 2137601. Available at SSRN: http://ssrn.com/abstract=2137601

[11] Directive 2009/136/EC