FACEBOOK TRACKS AND TRACES EVERYONE: LIKE THIS!

Arnold Roosendaal[*]

Abstract

Numerous websites have implemented the Facebook Like button to let Facebook members share their interests, therewith promoting websites or news items. It is, thus, an important business tool for content providers. However, this article shows that the tool is also used to place cookies on the user's computer, regardless whether a user actually uses the button when visiting a website. As an alternative business model this allows Facebook to track and trace users and to process their data. It appears that non-Facebook members can also be traced via the Like button. This means that Facebook's tentacles reach far beyond their own platform and members. Due to the extensive web coverage with Like buttons, Facebook has a potential connection with all web users. Web activity can be linked to individual accounts or a separate data set can be created for individuals who are not (yet) a Facebook member. The hidden collection of data on browsing behavior and the creation of individual data sets has implications for the privacy of individuals. This article discusses privacy issues arising from third party cookie use and connectivity of web activity and devices, using the technical process behind the Facebook Like button as an example.

Note: This paper is work in progress. A final version will appear later on.

[*] Doctoral Candidate and Researcher, Tilburg Institute for Law, Technology, and Society (TILT) Tilburg University, The Netherlands.

1. Introduction

The use of cookies and third party cookies to recognize and track and trace web users is not a new concept. Usually, the cookies are placed on the user's web browser without any visibility. In order to let third parties place cookies they have to be allowed to place content on a website. The content is requested from the web server of the third party and is delivered along with a cookie. When a site is visited again, the cookie is sent along in the request for the content. This allows content providers to 'remember' preferences of web users, such as language settings or purchasing history, and to provide the web content according to these preferences.

Tracking and tracing users over the web is a valuable tool for profiling purposes. Based on revealed interests web users can be targeted for personalized advertisements. Companies that earn their revenues from targeted advertising have a huge interest in using these techniques. It is, thus, not surprising that the way these techniques are exploited become more and more sophisticated. Sophistication can also be in the presentation. For instance, Facebook offers content providers to place a Like button on their web site. This button is a tool which allows Facebook members to indicate that they like a certain web site or item on a web site. By clicking the button, a link to the item is placed on their Facebook profile page. In addition, the number of visitors who 'liked' something is indicated next to the button. For content providers, the Like-button can, thus, function as an important business tool. Visitors contribute to attracting more visitors to a web site. This makes the tool valuable for content providers, which is also reflected by the fast increase in web coverage of the Like button. However, even though presented as a nice feature for content providers, the Like button is also used to place cookies and to track and trace web users, regardless of whether they actually use the button. The browsing behavior of individuals can be connected to their Facebook account. When a user has no Facebook account, a separate set of data concerning individual browsing behavior can be created. When a user creates an account later on, the data can be connected to the newly established profile page.

In this article, first a brief introduction to the Facebook Like button will be given (2). Then, in section 3 the technical process of placing and replacing cookies with the help of the button will be described, as well as how this facilitates profiling. Subsequently, the way this practice affects the privacy of individuals will be discussed (4) and, finally, a conclusion will be drawn (5).

2. The Facebook Like button

The Facebook Like button is an image displaying a thumbs-up symbol accompanied by the word 'Like'. According to Facebook, "[t]he Like button lets a user share your content with friends on Facebook. When the user clicks the Like button on your site, a story appears in the user's friends' News Feed with a link back to your website."[1] Anyone can implement the button on his website by simply adding the code which is available for free. The button can, thus, be used by content providers to have web users promote content and create links on their Facebook profile pages. When

[1] Facebook Developers, "Like Button" (2010) available at http://developers.facebook.com/docs/reference/plugins/like (accessed 26 Nov 10).

clicking the Like button, a login field opens in a pop-up window to login to Facebook. Logging in results in the creation of the link on the Facebook profile page. When a user is already logged in to Facebook the creation takes place immediately.

In April 2010, at their f8 conference, Facebook announced Instant Personalizer and Social Plug-ins, two services that allowed partners to leverage the social graph — the information about one's relationships on the site that the user makes available to the system — and provide a channel for sharing information between Facebook and third parties. For example, Web sites could implement a Like button on their own pages that enables users to share content from that site with the user's connections on Facebook.[2] The value of implementing the Like button on a website becomes clear from the statistics. Sites that added social plug-ins from Facebook, of which the Like button is one and is recommended to start with, reported increases in traffic of plus 200% and even more. Besides, the time spent and the number of articles read on websites with Like buttons also increased by over 80%.[3] The button represents 12.9% of the distribution of third party widgets.[4] It also appears that within months the use of social plug-ins has reached millions of sites.[5] The penetration rate of the Like button in the top 10.000 web sites reached over 4% in the first six months after its introduction[6], and it is likely that the penetration rate will continue to grow exponentially.

While the Like button can help content providers to generate traffic to their websites, it is also a tool for Facebook members to add information about their interests to their personal profile page. Therewith, it fits perfectly in the ongoing trend of social networking sites like Facebook to share personal information. Obviously, for sharing items from the web it is a very useful tool, for it allows direct linking without having to copy and paste complete URLs and the content is made up in a readable manner automatically.

3. Cookies, recognition, and identification

As indicated, there are numerous third parties which deliver content to websites and place cookies. Usually, the function of these third parties is to provide website providers with information on the number of visitors and which items on a website attracted the most attention. The third parties, thus, also provide a service to the website provider. These services are directly provided by the third parties what also implies that they have to receive the information on the visitors directly. This is facilitated automatically, because a piece of content is delivered from the servers of the third party and can be sent together with the cookie. A web user is usually not

---

[2] d boyd and E Hargittai, "Facebook Privacy Settings: Who Cares?" (2010) 15 First Monday 8.

[3] Facebook Media, "The Value of a liker" (2010) available at http://www.facebook.com/notes/facebook-media/value-of-a-liker/150630338305797 (accessed 26 Nov 10).

[4] BuiltWith, "Facebook Like Usage Statistics" (2010) available at http://trends.builtwith.com/widgets/Facebook-Like (accessed 26 Nov 10).

[5] J Constine, "Facebook Says "Likers" Click Links To External Websites 5.4x More" (2010) available at http://www.insidefacebook.com/2010/09/29/facebook-stats-likers/ (accessed 26 Nov 10).

[6] BuiltWith, "Facebook Like Usage Statistics" (2010) available at http://trends.builtwith.com/widgets/Facebook-Like (accessed 26 Nov 10).

aware of this. He just types in the URL of the website he wants to visit and the page is loaded. That the loading of the page involves numerous HTTP requests for content from the servers of the visited websites and often several third party servers is a process which takes place behind the scenes. More popular: that is where the magic happens!

A cookie is placed on the web users' computer via his browser. Only the server from which the cookie was sent has access to the cookie, so each cookie is connected to a web server. It is not the case that the provider of a website has access to all cookies placed by third parties via his website. Once a cookie is available on the user's computer, this cookie will be sent together with the HTTP request in each later request for content from the server which installed the cookie. The HTTP request also includes data on the referrer, which is the website on which the content will be displayed. Since the referrer data is always included, third parties can follow exactly which sites a user has visited and when. The content is needed when loading the page, so for being followed it is irrelevant whether a user actually clicks a piece of content or not.

Now take the Facebook like button. This is also a piece of third party content. It is not that the website provider directly places an image of this button on his website. In fact, the button is a piece of HTML code which includes the request to the Facebook server to provide the image when the website is loaded. This implies that the button can be used to set third party cookies or to recognize them as well. A few different scenarios can be distinguished. The scenarios have been tested in a practical experiment using Techcrunch.com, CNN.com, and Gizmodo.com.

3.1 The web user has a Facebook account

The first option is a scenario in which the web user has a Facebook account. When the account is created, Facebook issues a cookie containing a unique user ID. This cookie facilitates the display of a username in the login field at returning visits. When accessing Facebook from another device, a temporary cookie is issued, which is replaced by a cookie with the same ID after logging in to the account. This way, different devices can be connected to one account, and thus one user, by carrying the same ID cookie. Every time the user wants to visit the Facebook web site, the cookie is sent together with the HTTP request for the site. As a result, Facebook already knows who wants to log in before the actual login has taken place.

However, the cookie is not only sent to the Facebook servers when a member wants to log on, but in every occasion where content, such as the Like button, has to be provided from the Facebook servers (fig. 1). Thus, every single time a web site which includes the Like button is visited Facebook receives the information concerning the user, including his unique ID, via the cookie. When the user actually clicks the button, he has to provide his Facebook login details and a message about the 'Like' is posted on his profile page.

Since data about the user are sent to Facebook regardless of whether the Like button is actually clicked upon, users are often not aware of this fact. Nevertheless, the cookie contains the unique user ID and therewith facilitates the information on browsing behavior to be connected to the account. Even though the user is not involved, Facebook can collect far more individual data then the data made available on the profile page only.

```
GET
/plugins/like.php?href=http%3A%2F%2Fwww.facebook.com%2FGizmodo&layout=button_count&show_faces=f
alse&width=200&action=like&colorscheme=light&height=21 HTTP/1.1
Host: www.facebook.com
User-Agent: Mozilla/5.0 (Windows; U; Windows NT 5.1; en-GB; rv:1.9.2.10) Gecko/20100914 Firefox/3.6.10
Accept: text/html,application/xhtml+xml,application/xml;q=0.9,*/*;q=0.8
Accept-Language: en-gb,en;q=0.5
Accept-Encoding: gzip,deflate
Accept-Charset: ISO-8859-1,utf-8;q=0.7,*;q=0.7
Keep-Alive: 115
Connection: keep-alive
Referer: http://gizmodo.com/
Cookie: datr=yjPATCXPQuDBLU_J5ZfRsJpd; lu=TgbyaYN2Obo-F4fEBiQTGtwQ; locale=en_GB; x-
referer=http%3A%2F%2Fwww.facebook.com%2Fhome.php%23%2Fhome.php; cur_max_lag=20;
c_user=100001XXXXXXXXX; sct=1287731574; sid=0; xs=55dcbdfe4719c2693d477d0c0dd83ab6
Cache-Control: max-age=0
```

Fig. 1: The HTTP GET request for the Like button on Gizmodo.com, including the cookie with user ID (anonymised by the author with XXXXX)

In this scenario, there is a link between the Internet user and Facebook, because there is an account. Now, let's consider a scenario where there is no link.

3.2. The web user does not have a Facebook account

When a user does not have a Facebook account, there is no cookie and no user ID available. In this case, a visit to Techcrunch.com includes an HTTP GET request for the Like button. However, when the button is provided there is no cookie issued. Thus, it seems that the Like button itself is not used to issue cookies. However, when a site is visited which includes Facebook Connect (for instance Gizmodo.com) this application issues a cookie (fig. 2). From that moment on, visits to other websites which display the Like button result in a request for the Like button from the Facebook server including the cookie. An important part of the process depends on visiting a site which has implemented Facebook Connect. The chance of visiting such a site is considerable. Within a year from its launch in December 2008, Facebook Connect was used at almost 1 million web sites and in March 2009 over 40 million unique visitors of Facebook Connect implementations were registered.[7] The number of implementations increases exponentially, so the likelihood of passing by such a web site is becoming bigger at a fast pace as well.

As indicated, after visiting a web site on which Facebook Connect has been implemented, the request for the Like button includes a cookie. This cookie has an expiration date two years from the moment it was issued. However, by browsing across web sites, additional cookies can be placed on the user's computer and these can be added later on in new requests. Not all cookies are used in this way. For instance, a cookie issued via the extern login status plug-in is not included in later requests.

---

[7] K Burbary, "Five Reasons Companies Should be Integrating Social Media with Facebook Connect" (2009) available at http://www.kenburbary.com/2009/08/five-reasons-companies-should-be-integrating-social-media-with-facebook-connect/ (accessed 26 Nov 10).

Based on the cookie, the entire web behavior of an individual user can be followed. Every site that includes some kind of Facebook content will initiate an interaction with the Facebook servers, therewith disclosing information about the visited web site together with the cookie.

1. Set-Cookie: datr=ckviTDm3989eNbvw6xMhAWle; expires=Thu, 15-Nov-2012 09:14:26 GMT; path=/; domain=.facebook.com
2. Set-Cookie: datr=ckviTC8tNJ-1ZKqCu_SrIga7; expires=Thu, 15-Nov-2012 09:14:26 GMT; path=/; domain=.facebook.com

Fig. 2. A cookie issued via Facebook extern login status (1) and one via Facebook Connect (2) on Gizmodo.com.

3.3. The web user becomes a Facebook member

It is possible that a web user already has a personal set of data collected by Facebook, based on the mechanism described above. The question is what happens when this user creates a Facebook account. In that case, he first has to go to the Facebook homepage (login page). The cookie the user has on his computer is sent to Facebook in the request for the web page to be loaded. The server responds and issues a few new cookies. These new cookies are temporary cookies, or session cookies. When the account is actually created, a unique ID number is issued and sent in a cookie. The connection between this ID cookie and the old cookie is made behind the scenes by Facebook's servers. This means that the entire historical information of the user can be connected to the newly created Facebook account. From this moment on, all subsequent requests for Facebook content go accompanied with the cookie including the unique user ID.

When all cookies are deleted, the process starts from the beginning again with Facebook Connect placing a new cookie when a site is visited where Facebook Connect is implemented. However, from the moment on that an individual accesses his Facebook account, or connects to this account by clicking the Like button and providing username and password, this cookie is replaced by a cookie containing the unique user ID that belongs to the account.

The cookies are used for recognition. Web users can be recognized whenever they visit a site with a piece of Facebook content. Facebook members are identified as individual account holders, because the cookie includes their unique user identification number. When different devices are used to access Facebook, such as a home computer, a laptop, and a smart phone, these devices are recognized as belonging all to the same individual. So, all web interaction from these different devices is connected as well. Individuals who do not have a Facebook account are recognized as well. Their browsing behaviour is, however, not connected to a Facebook account. Besides, recognition is machine based and separated for every single device. Since there is no unique user ID in the cookie resulting from a log on to Facebook, the different devices cannot be connected solely on the basis of the cookies. Single devices can be quite reliable, however, even though they can be used by different persons. More and more devices, such as laptops and smart phones, become personal and are usually used by one single individual. This implies that information collected based on the cookies and browsing behaviour results in a very personal profile. Obviously, Facebook can use this to serve their members targeted advertisements. Most probable, the information collected about the browsing

behaviour of non-members can be used to have a larger sample for profiling and targeting purposes.

The Facebook Like button is not the only button which frequently appears on web sites to facilitate sharing or promoting of content. Other examples are Twitter's Tweet button, the Digg button, and Google's Buzz. There are, however, some differences. As described above, Facebook Connect is the system that actually issues a cookie the first time. From that moment on, the cookie is sent together with all HTTP requests for content, so also when the Like button has to be loaded on a page. Thus, an additional system is used to initiate the cookie exchange. Twitter, for instance, does not have such a system. The Tweet button does not always sent a cookie when the button is requested from the Twitter servers. This is only the case when someone has visited the Twitter homepage; then a cookie is issued which is used in future interactions with the servers, similarly as with the Like button. Logging in or even having an account at Twitter is not necessary. A (small but important) difference with the Like button is that there can at least be supposed to be some link to Twitter, because the web user has visited this web site. For Facebook, this is not necessary at all. This implies that individuals who consciously choose not to participate in Facebook are still tracked and traced by Facebook. When someone does not connect to Facebook himself, Facebook makes the connection.

Another important difference is that Facebook can connect the browsing behaviour to member accounts. These accounts are, usually, quite rich concerning disclosed information, but the Like button as exploited by Facebook makes that far more information is collected about individual members then the information disclosed on the personal profile page. Thus, people who have an account, but do not want to disclose that much information are still profiled more extensively. Their browsing behavior discloses much information concerning personal interests, and this information can also be collected by Facebook and connected to the individual account. In the end, consciousness in disclosing information, either by not participating on Facebook or by very limited disclosure of personal information, is not sufficient to escape Facebook's tentacles.

4. Privacy implications

The way the Facebook Like button is used to collect information concerning browsing behaviour of individuals clearly has implications for privacy. Even though it is difficult to give a clear-cut definition of privacy, notwithstanding the fact that various attempts have been made to describe the concept[8], some aspects are broadly recognized as essential in this respect. The two most prominent aspects are informational self-determination and contextual integrity.[9] These are reflected in regulations concerning personal data protection by means of requirements such as data minimisation, purpose specification, informed consent of the data subject, and data subject access rights. The aim of these requirements is to limit the processing of personal data to the least necessary and to provide the individual with some instruments to control the disclosure and use of personal data. In relation to

[8] For instance DJ Solove, "Conceptualizing Privacy" (2002) 90 California law review 4 1087; DJ Solove, "A Taxonomy of Privacy" (2006) 154 University of Pennsylvania law review 3 and; WA Parent, "Privacy, Morality, and the Law" (1983) 12 Philosophy and Public Affairs 4 269.

[9] H Nissenbaum, "Privacy as Contextual Integrity" (2004) 79 Washington Law Review 119.

informational self-determination the individual should be able to decide which data are disclosed to whom and for what purpose. The aspect of contextual integrity means that data have to be treated according to the norms applicable to the context in which the data were disclosed. Besides, data should not be transferred to another context without the individual's consent.

Now, when looking at the situation of the Facebook Like button, these basic requirements are infringed upon. First of all, the data collection takes place without the individual web users being aware. As a result, there cannot be consent for the data collection. Possibly, Facebook members have agreed to this by accepting the general terms and conditions when signing up for the social network site. Non-members, however, cannot have agreed to this and are subjected to the collection by merely using the internet. Second, the exact purposes for the collection of the data are not clear and the limitations are undefined as well. It can be expected that Facebook uses the data for targeted advertisements on their web site. The data of the individuals who are not a member can be used to have a bigger sample. Nevertheless, these individuals cannot be subjected to the advertisements themselves, since they are not visiting the Facebook web site. Finally, data subjects cannot use their access rights to review the data and ask for correction or deletion.

In the argument as set out above, the data were considered to be personal data in the sense of data protection legislation. For the data related to Facebook members this can easily be defended. The unique user identification number belonging to the personal profile of the member is included in the cookie and sent to the Facebook servers in the HTTP request.[10] Even when Facebook would state not to connect the data to the individual profile, the data have to be considered personal data, because for determining whether a party is able to link the data to an identifiable individual all means available to that party have to be taken into account. Facebook clearly has the user data and identification numbers, so they can link the data to an identifiable individual.

For the data belonging to web users who do not have a Facebook account, the argument is a bit more difficult. However, an extensive data set can contain a huge amount of information which can clearly indicate who the individual is, or at least create a detailed picture of individual preferences and interests. As a result, data sets become more and more personal and facilitate individualisation. Besides, it has to be taken into account that devices become more and more personal. With the increasing use of laptops and smart phones devices are not as much shared by different users as was the case a couple of years ago when the personal computer was serving entire families. When a web user later decides to create a Facebook account the data are linked to that account and are certainly personal data.

A more general concern is the process behind the Like button and the way the button is presented. While it is offered as a tool for web site owners to implement social features and therewith attract more visitors, there is a big interest for Facebook for having the button implemented in as many sites as possible. Whether the button is actually used by its members is of minor concern, because data concerning browsing

[10] This in contrast with what was argued by a Facebook spokeswoman who said that identification was based on the IP address and browser of the web user (OUT-LAW News, "NHS criticized for sharing website data with Facebook" (2010) available at http://www.out-law.com/page-11576 (accessed 26 Nov 10)). However, as can be seen in Fig. 1 the ID is included. It is just not that the original user-ID cookie is sent along.

behaviour are communicated to Facebook anyway. However, the most prominent concern is that web users are somehow mislead. Due to the way the button is presented, web users do expect to have data transferred when they use the button. That data are transferred even when the button is not clicked upon is difficult to imagine for the ordinary web user. In addition, web users who have no connection with Facebook at all do certainly not expect to have their data transferred to this social network site. As a result, individuals who consciously choose not to participate in the social network site are still connected to this web site.

5. Conclusion

In this article, the Facebook Like button was discussed as an example of a third party cookie which is exploited in a smart way. The button is presented, and valuable, as a business tool for web site holders and content providers. However, the button is used to collect detailed data about browsing behaviour of individual web users. These data can be connected to the Facebook profile page or collected as a separate data set. By using cookies and unique identification numbers, Facebook has a potential connection with every web user and can track and trace individual web behaviour. This practice raises privacy concerns and conflicts with informational self-determination of individual web users. Facebook tracks and traces everyone; Like this!