Online Tracking, Targeted Advertising and User Privacy - The Technical Part

Privacy and Web 2.0 Seminar Summer Term 2011

Simon Sprankel
sprankel[at]rbg.informatik.tu-darmstadt.de

September 30th, 2011

This paper gives a short overview over the field of Online Tracking, Online Targeted Advertising (OTA) and the related privacy issues. Balachander Krishnamurthy rightly states that “there are at least three different angles through which one could approach the problem of reducing privacy leakage: technical, legislative and economic” [1]. This paper focuses on the technical part, whereas Nadine Trüschler’s paper mainly focuses on the legislative and economic part. Since she also gives a more detailed introduction in the topic of OTA, it is advantageous to read her paper first.

1 Introduction

A short analysis of network traffic suffices in order to see that tracking of users over multiple websites is an up-to-date topic and raises various privacy concerns. For instance studiVZ, one of the largest German *Online Social Networks (OSNs)*, communicates ones profile ID to a big advertising company, Tradedoubler. Another example is a Facebook app called Kickmania, which communicates ones profile ID to another advertising company [2]. Hence, these companies are able to identify the user via its public profile at the OSN. Data, which makes a person identifiable is called *Personally Identifiable Information (PII)*. There are numerous other cases in which web applications delegate PII to advertising companies. With the growing use of online social networks, the problem has even become worse. It has been shown, that “it is possible for third-parties to link PII, which is leaked via OSNs, with user actions both within OSN sites and elsewhere on non-OSN sites” [2]. Thus, it is an interesting task to look at the ways users are tracked and the ways this can be prohibited.

2 Web Tracking Technologies

Web tracking means that a user is tracked over various websites by e. g. advertising companies. If an advertising company is able to track a user, it can create an interest profile of the user. This process is often referred to as *profiling*. With an interest profile, it is possible to show *targeted advertisements* to the user. This is also called *behavioural targeting*. In general, these advertisements are more interesting to the user, so that he will click the targeted advertisement more often than a random one. Studies claim that without targeted advertising, advertising effectiveness decreases by around 65 percent [3]. Hence, advertising companies have a huge interest in web tracking. Anyway, the user's privacy must not be forgotten.

There are innumerable types of web tracking technologies. I will focus on cookies, web bugs and fingerprinting. There are also other possibilities like browser extensions, complex JavaScript code (e. g. used by Google Analytics), modified browsers or deep packet inspection.

2.1 Cookies

Cookies are arbitrary strings belonging to a website and are stored on the user's machine. Their intent is to "facilitate a browser-server stateful interaction, in a stateless protocol" [4]. Each time a user contacts a website, the server may send cookies to the client, which are then stored by the user's browser. If a user contacts the same website again, the stored cookies are added to the request sent to the server. So it is possible to store user-specific data, e. g. shopping cart information or language preferences, in cookies. As one can see, cookies may be used for quite reasonable things.

However since websites include more and more third-party content, not only cookies of the requested website, but also many third-party cookies are set. The user often does not expect to communicate with another company than the one the requested website belongs to, so these third-party cookies raise several privacy problems. They are often set by advertising companies. The advertising companies place advertisements on many websites and as a result, they are able to track the user over these websites.

Technically, there are different types of cookies. I will introduce HTTP cookies, flash cookies and evercookies.

HTTP cookies are the commonly known cookies stored in the user's browser. They are set by the Set-Cookie header described in RFC 6265 [5] and easy to delete in the menu of all current browsers.

Flash cookies (also called Local Shared Objects LSO) are somehow more advanced cookies, since they are more difficult to delete. They can be set from a flash object integrated into a website. They are not stored in a folder controlled by the browser, but in a folder configured by Adobe. In the Linux distribution Ubuntu, the folder is */home/USER/.macromedia/Flash_Player/#SharedObjects*. Since Flash 10.3, Adobe provides the so called ClearSiteData-API, with which it is possible to delete flash cookies.

The current versions of Internet Explorer, Firefox, Opera and Google Chrome already support the deletion of flash cookies. Solely Apple's Safari lacks such a feature by birth. Just a year ago, before Adobe released the ClearSiteData-API, it was only possible to delete flash cookies manually or on a specific website of Adobe.

Evercookie [1] is a JavaScript API enabling websites to create cookies that are nearly impossible to delete. Currently, evercookie combines 13 different types of cookie storage possibilities like HTTP cookies, flash cookies, Silverlight and HTML5 storage functions, history and cache techniques and others. If data like HTTP and flash cookies are removed, the data can easily be restored, because it is saved redundantly.

Techniques to make cookies permanent, often referred to as *cookie respawning*, have been used in the wild by e. g. KISSmetrics [2]. Ever since some researchers at U.C. Berkeley found this out [11], the topic is in the press. Now, there are first lawsuits against cookie respawning methods [3], so the end of these techniques may be near.

2.2 Web Bugs

Web Bugs, also called web beacons, tracking bugs or page tags, are "1x1-pixel pieces of code that allow advertisers to track customers remotely" [3]. In general, they are invisible to the user. The most common form of them are tracking pixel, which are 1x1 images often included from a third-party site. As an example, Google uses them to track conversions [4]. Here is an example code for Googles tracking pixel:

```
<img height="1" width="1" border="0" src="http://www.googleadservices.com/
pagead/conversion/1234567890/?value=100&label=Purchase&script=0">
```

With this pixel, a conversion ID (1234567890), the total amount of the order (100) and the type of the conversion (purchase) is transferred to Google. Of course, it is possible to transfer much more data in these tracking pixels.

2.3 Fingerprinting - Panopticlick

A web tracking technology that is nearly impossible to block is fingerprinting. A fingerprint is a summary of software and hardware settings. Fingerprints often identify a user uniquely, so that they may be used as a global identifier. In combination with the IP address of the user, a fingerprint could also be used as a cookie regenerator, if cookies are deleted. As a last usage scenario, fingerprints in combination with the IP address could also be used as an identifier, if cookies are completely disabled.

Fingerprinting is rather hard to detect because it does not leave any traces on the user's system. This is also the reason why it cannot be deleted or prohibited except by a configuration change making the system less unique.

[1] http://samy.pl/evercookie/

[2] http://www.kissmetrics.com/

[3] http://www.wired.com/epicenter/2011/08/tracking-lawsuit/

[4] If a user clicks on an advertisement and directly achieves a goal specified by the website operator, e. g. placing an order, this is called a conversion.

Panopticlick [5] is a research project of the Electronic Frontier Foundation studying fingerprinting based on browser uniqueness. In this project, Peter Eckersley learned that (in his sample data of $470,161$ participants) "83.6% of the browsers seen had an instantaneously unique fingerprint, and a further 5.3% had an anonymity set of size 2. Among visiting browsers that had either Adobe Flash or a Java Virtual Machine enabled, 94.2% exhibited instantaneously unique fingerprints and a further 4.8% had fingerprints that were seen exactly twice" [6]. These numbers show that fingerprinting is a scary approach and has the capability to become the major tracking technique.

3 Solutions To Prevent Tracking

As we already saw, there are many possibilities to track users and a lot of websites use one or more of these possibilities. This raises the question how one can prevent tracking. The most common solutions that are partly implemented, cookie blocking, domain blocking and Do Not Track, are described in this section.

3.1 Disable Cookies / Cookie Blocking

A simple solution to prevent tracking through cookies is to block them [7]. One can do this generally for all sites or site by site. To block them site by site is a huge effort - who wants to decide whether to allow cookies or not for every new page one visits? As a matter of fact, this solution is effective in order to prevent tracking through cookies. Anyhow, there are many other, more sophisticated possibilities to track users than cookies. Additionally, many sites require users to enable cookies and do not let them access the site if cookies are disabled. One more disadvantage is that auto login functions, also known as remember me functions, do not work without cookies.

This solution is supported by every current web browser.

3.2 Domain Blocking

Another common solution is to block any connection to websites of advertising companies [7]. This method is often used in ad block extensions like Adblock Plus [6]. For people who do not want to see any advertisements at all, this is a good solution. However, there are people who like to see advertisements but do not want to be tracked. In addition, it is generally difficult to maintain a list with domains which should be blocked. Adblock Plus for example depends on user input for these lists.

Microsoft proposed a format for such filter lists [8] and already implemented an ad blocker based on these filter lists in Internet Explorer 9 [7]. Unfortunately, one has to activate and set it up manually. The already mentioned Adblock Plus is available for

[5] https://panopticlick.eff.org/

[6] http://adblockplus.org/

[7] http://windows.microsoft.com/en-US/internet-explorer/products/ie-9/features/tracking-protection

Firefox and Google Chrome. AdBlock for Safari [8] blocks advertisements in Safari. An ad blocker for Opera is also available [9].

3.3 Do Not Track - An FTC Proposal

In December 2010, the Federal Trade Commission (FTC) proposed Do Not Track (DNT), "a uniform and comprehensive consumer choice mechanism for online behavioural advertising" [9]. The idea behind DNT is that the user is able to tell every website whether he wants to be tracked or not. This shall be possible by placing a persistent setting in the browser. The idea is not new at all - it is adapted from the US Do Not Call Registry [10], where people can register their telephone number so that they will not be disturbed by telemarketing calls.

There are different implementations of DNT, which all have more or less the same benefits and drawbacks. A major advantage of DNT is that it is partly implemented in recent browsers. Additionally, it is a good idea to let users tell websites whether they want to be tracked or not. Unfortunately, websites can ignore the user request not to be tracked. So the enforcement of this solution is a rather big and nearly unsolvable problem.

(Permanent) Opt-Out Cookies The idea of opt-out cookies is to store a cookie for each advertising company from which one does not want to be tracked [7]. The cookie is a simple one: its name is `DNT` and its value is `1`. It is a complex solution, since a new cookie is needed for every domain. Hence, one has to get support for creating them. On the website of the Network Advertising Initiative [11] and on a more general website [12], one can create the opt-out cookies. Since cookies may be deleted, one is also dependent on support for making these opt-out cookies permanent. Fortunately, there are browser extensions that can do both - create the cookies and make them permanent. These extensions normally check whether cookies are deleted. If this is the case, they directly create new opt-out cookies.

For Firefox, Targeted Advertising Cookie Opt-Out (TACO) [13] is available. For Chrome, it is Keep My Opt-Outs [14] or the more complete Keep MORE Opt Outs [15]. For Opera, Safari and Internet Explorer, it seems that no similar extensions exist.

Do Not Track DOM Property Another implementation of DNT is to create a new DOM property for it. DOM stands for Document Object Model and is a language-independent definition for representing and interacting with XML documents. If a user activates the DNT feature in his browser, the DOM property `document.navigator.doNotTrack`

[8] http://www.safariadblock.com/

[9] https://addons.opera.com/addons/extensions/details/opera-adblock/0.50/

[10] https://www.donotcall.gov/

[11] http://www.networkadvertising.org/managing/opt_out.asp

[12] http://www.aboutads.info/choices/

[13] https://addons.mozilla.org/en-US/firefox/addon/targeted-advertising-cookie-op/

[14] https://chrome.google.com/webstore/detail/hhnjdplhmcnkiecampfdgfjilccfpfoe

[15] https://chrome.google.com/webstore/detail/eoibfeagdaaoimfpfalgbmmegagdconp

`== "1"` has to return `TRUE`. The property could be queried in client-side code before tracking-related actions are taken [7]. Since only the client has access to the DOM, it is reasonable to use both, the DNT DOM property and the HTTP header explained below. Standardisation of the new DOM property has been proposed to the W3C [16] [8].

A benefit of this solution is that it may reduce server requests for users who opted out from tracking. If this property is activated, there should not be any server request to advertising companies at all, so that the user can check whether his preference is honoured. A drawback is that websites could force users to disable the property in order to access their content.

The DNT DOM property is only supported by Internet Explorer (with a vendor prefix: `navigator.msDoNotTrack`) and Safari.

Do Not Track HTTP Header The DNT HTTP header is another approach to implement DNT. Where the DOM property can only be used by client-side tracking scripts, the HTTP header tries to solve the DNT issue for server-side tracking scripts. The idea is to extend the HTTP header by a `DNT` field, whose value is supposed to be `1` if tracking should be disabled. So with every server request, the browser expresses the user's wish not to be tracked to the server. Then, the server may or may not respect this setting. Standardisation of the HTTP header extension has been proposed to the W3C [8]. There is also an internet-draft of the Network Working Group of the IETF regarding this HTTP header extension [10].

This solution is implemented in Firefox, Internet Explorer and Safari.

Compliance & Enforcement The biggest problem with all DNT solutions is to check whether the DNT setting is respected. It is not a solution that completely disables the possibility of web tracking. It is just an expression of the users preference and can therefore be ignored. Even if advertising companies pretend to respect the setting, nobody is able to check that. To be sure that the setting is respected, one would have to check the code on the servers of the advertising companies. Of course it is unrealistic that these companies will publish their code or grant access to their servers. Cooper and Tschofenig mentioned that "this sort of guarantee [17] may require enforcement or intervention from governmental privacy authorities in order to truly be effective" [7]. Hence, it is not only a technical problem, but more and more an interdisciplinary one.

3.4 Browser Support Overview

Table 1 shows an overview of how browsers support the just introduced solutions to prevent tracking. The current versions of the five most used browsers [18] have been used for testing.

[16] World Wide Web Consortium

[17] the guarantee that the DNT setting is respected

[18] `http://www.browser-statistik.de/statistiken/`

	Mozilla Firefox	Google Chrome	Microsoft Internet Explorer	Apple Safari	Opera
Disable Cookies / Cookie Blocking	supported	supported	supported	supported	supported
Domain Blocking	extension available	extension available	supported	extension available	extension available
Permanent Opt-Out Cookies	extension available	extension available	not supported	not supported	not supported
DNT DOM Property	not supported	not supported	supported	not supported	not supported
DNT HTTP Header	supported	not supported	supported	supported	not supported

Table 1: Overview of the browser support of the solutions to prevent tracking

3.5 Effectiveness Of The Solutions

As a matter of fact, even if all solutions to prevent tracking above are taken into account, web tracking is still possible. The first thing is that tracking is always possible if the first party gives PII to a tracking company. Even if one is able to block all third party contents on a page, the first party itself could still forward PII.

Additionally, there are no proper technical solutions to stop tracking via e. g. Evercookies (2.1) or Fingerprinting (2.3) yet. If it is not possible to block web tracking technologies, it may be a good idea to make web tracking needless. Exactly this is the approach of the next section.

4 Privacy Preserving Targeted Advertising

As stated in the sections above, it is not yet possible to stop web tracking and probably it will never be possible. Hence, another approach to “stop” web tracking has evolved - make web tracking needless. The biggest reason why companies track users is to profile them so that they are able to show targeted advertisements. If it is possible for these companies to use targeted advertising without tracking users, they do not have to use web tracking anymore.

So how can this be achieved? The most obvious possibility is to do the profiling on the users local machine and not on the server. The two most popular systems that use this technique, Privad and Adnostic, are introduced in this section. A more general system, RePRIV, which controls the access of websites to private user data, is also presented in this section.

Nevertheless, these systems do not prohibit web tracking, they only make it unnecessary.

4.1 Privad

Privad [19] is a non-tracking advertising system designed by researchers at the Max Planck Institute for Software Systems. It is implemented as a Firefox extension and can be downloaded from its website. It extends the standard advertising model with users, publishers and advertisers by two other entities - a broker and a dealer [12]. Advertisers upload their advertisements to the broker, so that the broker contains all advertisements. The dealer as an intermediary between the user and the broker makes the client anonymous by mixing messages from various users to the broker. There are two message types - one for sending the user interests to the broker and getting back the corresponding advertisements and one for sending the reports to the broker. The messages are all encrypted via a public key infrastructure, so that the dealer is not able to eavesdrop. Hence, in both cases, privacy is preserved: "The advertiser knows what is in the report, but not who sent it. [...] As with the reports, the intermediary cannot see what ads you receive, and the advertising company [20] does not know who got what ads." [19]. Although the dealer should be run by an independent and hence trustworthy organisation, the system architecture makes it possible that both, the broker and the dealer may be untrusted and privacy is still guaranteed.

Now why should the involving parties deploy Privad? For users, privacy is massively enhanced since profiling is no longer done on external servers, but locally on their machine and PII is not leaked to any other party. For advertising companies, it should be quite interesting, since the interest profiles of the users are much more fine-grained when built on the users machine. Additionally, they do not have to provide servers profiling the users, so that profit may be improved. As a last advantage, systems like Privad may improve performance because profiling on a local machine is much more efficient than on servers.

These advantages also apply to Adnostic, which is a quite similar system.

4.2 Adnostic

Adnostic [21] is also a privacy preserving targeted advertising system designed by researchers at the New York and Stanford University. A Firefox extension which implements this system is available as well. Since the system is quite similar to Privad, I

[19] `http://adresearch.mpi-sws.org/`

[20] in this case, this is the broker

[21] `http://crypto.stanford.edu/adnostic/`

will only have a quick look at the key components of it - profiling, ad insertion and accounting [13]. The profiling mechanism is similar to the one of Privad since it is also done locally in the users browser. The categorisation of visited pages is based on lists mapping URLs to their classification and natural language processing for URLs that are not included in this list. However, ad insertion is a bit different. If Adnostic is installed, a list of n advertisements matching the pages topic are downloaded and the browser extension than decides which advertisement fits best based on the user profile. Regarding accounting, it remains the same for the "cost per click" model, whereas accounting for the "cost per impression" model is done by a cryptographic protocol [22].

4.3 RePRIV

RePRIV [23] stands for Re-Envisioning In-Browser Privacy and is a system developed by Microsoft Research. As stated in the introduction, it is "a [more general] system for managing and controlling the release of private information from the browser" [14]. Until now, it is only developed on top of C3, a research browser developed in .NET [24]. In RePRIV, the browser mines user-related data and creates an interest profile. Service providers can register extensions that extract additional data from browsing activities called miners. For instance, a Twitter miner would be able to extract information from the Twitter activities of the user and insert it in the users interest profile. Every time a website requests data of the user, the user is asked by an explicit prompt whether specific data may be published to that provider. With this concept, users have the complete control over their personal data, but it may be annoying to answer the prompts repeatedly. This problem is mitigated in RePRIV by learning user preferences quickly and thus being able to answer the prompts automatically.

5 Conclusion

There are a lot of possibilities to track users over various websites; some of them have been introduced in this paper. Although there are already solutions to stop these tracking activities, some of them like Fingerprinting seem to be unstoppable and uncontrollable. Both, web tracking technologies and solutions to prevent them, develop and tracking technologies will probably always be one step further. Hence, the idea to make tracking needless by developing targeted advertising systems that preserve privacy seems to be forward-looking. However, the secure deployment of such systems demand cooperation of users, lawmakers, advertising companies and independent organisations. Lawmakers in cooperation with independent organisations have to enforce that the user is not tracked when such a system is installed.

The field of online tracking, targeted advertising and user privacy develops fast, so that this paper needs to be updated continually.

[22] The definition and explanation of the cryptographic protocol would exceed the scope of this paper.

[23] `https://research.microsoft.com/en-us/projects/repriv/`

[24] https://research.microsoft.com/apps/pubs/default.aspx?id=150010

Acknowledgements I would like to thank Hervais Simo Fhom and Christine Maul-Pfeifer for their feedback and comments.

References

[1] Balachander Krishnamurthy. *I know what you will do next summer (October 2010).* AT&T Labs-Research, `http://www2.research.att.com/~bala/papers/ccr10-priv.pdf`

[2] Balachander Krishnamurthy and Craig E. Wills. *On the Leakage of Personally Identifiable Information Via Online Social Networks (July 22, 2009).* `http://conferences.sigcomm.org/sigcomm/2009/workshops/wosn/papers/p7.pdf`

[3] Goldfarb, Avi and Tucker, Catherine. *Privacy Regulation and Online Advertising (August 4, 2010).* `http://ssrn.com/abstract=1600259`

[4] R. Tirtea, C. Castelluccia, D. Ikonomou. *Bittersweet cookies. Some security and privacy considerations (February 02, 2011).* `http://www.enisa.europa.eu/act/it/library/pp/cookies/`

[5] A. Barth. *HTTP State Management Mechanism (RFC 6265).* `http://www.ietf.org/rfc/rfc6265.txt`

[6] Peter Eckersley. *How Unique Is Your Web Browser?.* `https://panopticlick.eff.org/browser-uniqueness.pdf`

[7] A. Cooper, H. Tschofenig. *Overview of Universal Opt-Out Mechanisms for Web Tracking (March 07, 2011).* `http://tools.ietf.org/html/draft-cooper-web-tracking-opt-outs-00`

[8] A. Zeigler, A. Bateman, and E. Graff. *Web Tracking Protection (February 24, 2011).* `http://tools.ietf.org/html/draft-cooper-web-tracking-opt-outs-00`

[9] FTC Staff Report. *Protecting Consumer Privacy in an Era of Rapid Change.* `http://www.ftc.gov/os/2010/12/101201privacyreport.pdf`

[10] J. Mayer, A. Narayanan, and S. Stamm. *Do Not Track: A Universal Third-Party Web Tracking Opt Out (March 07, 2011).* `http://tools.ietf.org/pdf/draft-mayer-do-not-track-00.pdf`

[11] Mika D. Ayenson, Dietrich J. Wambach, Ashkan Soltani, Nathaniel Good & Chris Jay Hoofnagle. *Flash Cookies And Privacy II: Now with HTML5 and ETag Respawning.* `http://ssrn.com/abstract=1898390`

[12] Saikat Guha, Alexey Reznichenko, Kevin Tang, Hamed Haddadi, Paul Francis. *Serving Ads from localhost for Performance, Privacy, and Profit.* `http://conferences.sigcomm.org/hotnets/2009/papers/hotnets2009-final27.pdf`

[13] Vincent Toubiana, Arvind Narayanan, Dan Boneh, Helen Nissenbaum and Solon Barocas. *Adnostic: Privacy Preserving Targeted Advertising.* http://crypto.stanford.edu/adnostic/adnostic.pdf

[14] Matthew Fredrikson, Benjamin Livshits. *REPRIV: Re-Envisioning In-Browser Privacy.* http://research.microsoft.com/en-us/um/people/livshits/papers/tr/repriv_tr.pdf

[15] The Keynote Benchmark. *Tracking, Targeting & Trade-Offs.* http://www.keynote.com/benchmark/downloads/Browser_Privacy.pdf