Radboud University Nijmegen

Master Thesis Information Science

Research Number: 157 IK

Informed Consent in Behavioral Advertising

Author:
Philipp van Bebber
(0608785)

Supervisors:
Prof. mr. Dr. M. (Mireille) Hildebrandt
Dr.ir. E. (Erik) Poll

October 11, 2011

Preface

First of all I would like to thank my supervisors Mireille and Erik for all their support and guidance. We had numerous feedback meetings and discussions which I experienced as very helpful and interesting.

Mireille helped me understand the complex world of EU and national legislation. Erik helped me understand the technical background of behavioral advertising. Both were always available for all my questions.

Finally, I would like to thank my family and friends that supported me throughout my study.

Contents

Chapter 1

Introduction

1.1 Problem Statement

Nowadays, a lot of companies use the internet to attract new customers. The online advertising market is still growing and publishers turn much attention to behavioral advertising. We talk about behavioral advertising when the online behavior of an individual determines which advertisements will be presented. Different advertising methods such as e-mail marketing, social network advertising or banner ads, are used to get in contact with potential customers. Most of the advertisements that are presented to the user are well selected. Not every user gets to see the same advertisements when he or she visits the same website. Internet publishers and advertisers create user profiles of individuals who visit a website that contains advertisements or other embedded third-party content such as a YouTube video. Embedded content makes it easy for advertisement publishers to track the behavior of an individual [Con08]. Creating a profile can be done in a lot of different ways. One of the most common ways is to place cookies on the computer of an individual. Every time when the individual visits a website with embedded content the advertising publisher gets more information about the online behavior of that individual. This can also be achieved by using the IP as an identifier of an individual but cookies are the preferred ones[1]. Publishers offer to webmasters analysis tools and advertising profit. The only thing they want in exchange is that webmasters add a few lines of HTML or JavaScript to their pages [Con08]. The reason why most companies turn their attention on behavioral advertising is that they believe that users who belong to a certain group (profile) are most likely to be interested. Advertising companies will always mention the positive effects of behavioral advertising. Individuals get advertisements that match their interests better

[1] http://www.nytimes.com/2008/06/29/business/worldbusiness/29iht-ad30.1.14068922.html (accessed 25-04-11)

but there are also many privacy related concerns. One of the most important concerns is that individuals will be treated different because their behavior matches with a certain profile. This can influence the life of an individual in a positive but also in a negative way. Another concern is the privacy issue itself. Nobody, except the advertising company, knows what kind of data an advertising company possesses. Is it anonymous data or can it be linked to a real person? What happens in case that the data is published on the internet? Is it legal to collect behavioral data of an individual? Is consent necessary? These are questions that could be asked by everyone who accesses the internet. Especially the question about consent gets a lot of attention. A survey in 2009 [TKH$^+$09] carried out that 66% of adult US citizens do not want to be tracked by advertising companies. But how can we achieve that individuals can decide whether or not to get profiled by advertising companies? Can this be done in a way that individuals will still have access to the information they are looking for?

The European Union enacted two directives that deal with privacy and personal data:

1. Data Protection Directive (95/46/EC)
2. ePrivacy Directive (2002/58/EC)

Both directives say that *informed consent* is needed before personal data may be collected. In a nutshell, individuals must be aware of what kind of personal data will be collected, which methods will be used for data processing and for what purpose the data will be collected. European and national law do not always give a clear guidance on how to apply the articles of the directives. In which case can we talk about *informed consent*?

Also the Council of Europe made some recommendations on how to deal with profiling and personal data [oE10]. There exists uncertainty regarding the applicability of the Data Protection Directive with regard to profiling [SHKV08] because the directive is not applicable in case of anonymous data.

A lot of research on privacy enhancing technologies (PETs) is done in the last years. A list with many PETs exists from The Stanford Center for Internet and Society[2]. There are also EU funded projects such as Prime, FIDIS and PrimeLife that deal with identity management in Europe. We have to mention that these technologies focus on blocking cookies, banners etc. and do not help users to understand what is happening with their data. Thus, there is a need for transparency enhancing technologies (TETs) that do not only provide information about what data is gathered, how the data is gathered, by whom and why but also provide access to or information

[2] `http://cyberlaw.stanford.edu/wiki/index.php/PET` (accessed 23-05-11)

about the profiles someone matches. The information that is presented by a TET to the user must be easy to understand to guarantee that *informed consent* is given. Without a clue as to what profiles a person matches, there is no *informed consent.*

Research Question
How can we achieve informed consent in behavioral advertising profiling?

1.2 Research Method

This master thesis consists of five chapters. In the first chapter we will give a short introduction and discuss research domain, problem statement and research method.

The second chapter deals with research on behavioral advertising. When do we talk about behavioral advertising? Which methods are used to generate a profile? Furthermore, we will describe how advertising networks, websites, social networks and ISPs track their users.

Chapter 3 deals with applicable law with regard to privacy and personal data protection. The following questions are discussed: Which law or directive is applicable in case of profiling? How do advertisers achieve informed consent? Are there any gray areas in law? What kinds of privacy measurements are taken?

Chapter 4 deals with research on already existing methods like third-party cookie blocking, Do Not Track, Track Me Not, Tracking-Protection-Lists and P3P. What do the methods actually effect? How do they guarantee informed consent? Are users aware of how their data will be processed? Finally, it is necessary to determine if the discussed privacy-enhancing technologies can achieve *informed consent.*

Chapter 5 deals with analyzing the results and shortcomings of chapter 3 and 4. Finally, we will give a concept and recommendations about how to deal with the problem of achieving *informed consent.*

Chapter 2

Background

In this chapter, first a general overview is given. In the following section, the concept of behavioral advertising is explained. Finally, stakeholders, tracking methods and location-based services are discussed.

2.1 Overview

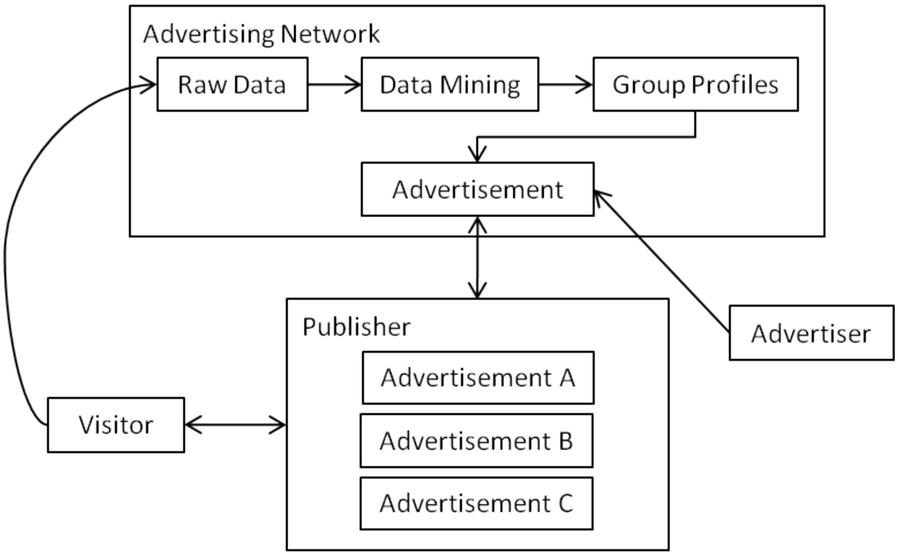


Figure 2.1: Behavioral Advertising overview

At the beginning, a company (advertiser) wants to publish advertisements about the products/services they offer on the internet. The advertisements are delivered to the advertising network and the company pays the advertising network for the publication of advertisements. Every time a visitor visits a website or search engine of a publisher information about the visitor is sent to the corresponding advertising network (see figure 2.1). This information is gathered by an advertising network and we call it raw data. Advertising networks apply data mining algorithms to the raw data and generate group profiles. These profiles are then added to relevant advertisements. Visitors will only see advertisements that belong to their group profile. It is possible that the advertising network, the advertiser and the publisher belong to the same organization. Google, for example, incorporates all three stakeholders, advertising network, advertiser and publisher.

2.2 Defining Behavioral Advertising

In this section we will describe how behavioral advertising works. Data mining as an ongoing process is described in section 2.2.1. By applying data mining techniques we can generate group profiles (section 2.2.2) that have certain characteristics.

We talk about behavioral advertising or behavioral targeting when information about an individual's online behavior is gathered and analyzed in order to show advertisements that statistically match with the interests of that individual. A lot of similar definitions of the term behavioral advertising/targeting exist. We will use the following one:

> *Behavioral targeting customizes messages to individual consumers based on their specific shopping interests, and characteristics like gender, age, and ethnicity. Behavioral targeting is a generic name for a series of technologies that collect and organize click stream data, develop data warehousing structures, apply data mining algorithms to uncover consumer browsing patterns, and serve targeted ads matched to an individual [Dwy09].*

One shortcoming of this definition is that the term "profile" is not used. Before customized advertisements can be presented to an individual we first have to categorize behavior and interests. Thus, a profile is created/applied. We have to distinguish between individual profiling and group profiling. In case of behavioral advertising group profiling is used for categorizing individuals. Individual profiling is applied when data of a single individual is used to construct a profile [CO08]. Also unique identification can be achieved by individual profiling which raises problems with regard to the topic "privacy" [Cus04].

Furthermore, two additional advertising methods exist: contextual advertising and demographic advertising. In case of contextual advertising, the content of a website determines the content of an advertisement. For example, a websites of a travel agency could publish advertisements about beach wear. Demographic advertising focuses on different consumer groups. Characteristics such as education, age or sex are used to allocate consumers to a group.

2.2.1 Data Mining

According to the Cross Industry Standard Process for Data Mining[1] a data mining project is a staged process which can be repeated. The process is illustrated in figure 2.2.

1. Business Understanding
During the first phase information about the project objectives and requirements are gathered. This knowledge will be used to describe the data mining problem and how to achieve the goal. In our case, advertising networks have to understand the purpose of an advertisement. Does it offer a product or does in turn attention on getting environmentally aware?

2. Data Understanding
First, data has to be collected. According to Custer [Cus04] this can be done in an explicit or implicit way by asking people or by observation of their behavior. It is also possible to use already existing data. Advertising companies have huge databases that they use for data mining purposes. After data is collected, first concepts can be discovered and subsets can be created.In terms of behavioral advertising, tracking information (see section 2.4) and information that is delivered by an individual itself is collected.

3. Data Preparation
The final data set is constructed by rearranging and ordering the raw data set. Also initial grouping of data items is sometimes useful. For example, with regard to behavioral advertising, events such as banner clicks can be grouped when they fall into a certain amount of time. Sometimes it is even better to do this during data collection [Cus04]. Further, it is necessary to have a closer look at the data mining algorithms that will be used in the next phase. A data mining algorithm cannot handle all data sets. Most of them require preconditions that have to be met. Thus, the data set has to be "cleaned" first.

4. Data Mining (Modeling)
Now, different data mining algorithms are applied. Parameters are often changed to get better results that fulfill the needs. It is possible to compare the results of different algorithms and to check whether an algorithm

[1] `http://www.crisp-dm.org/Process/index.htm` (accessed 05-05-11)

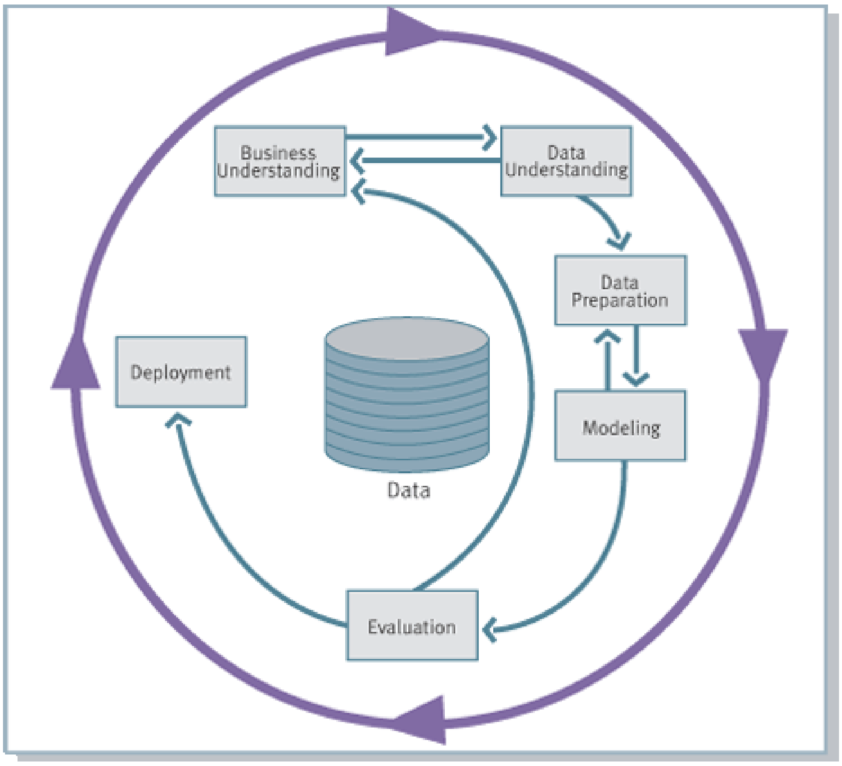


Figure 2.2: Phases of the CRISP-DM Process Model[2]

scores better on certain items of interest. Custer [Cus04] mentions one big difference between statistical methods and data mining techniques. In case of data mining it is not necessary to know what is being looked for. The strength of data mining is that it can deliver new, unknown, relations or patterns. Presumptions can be confirmed but also be falsified. With regard to group profiling clustering, classification and regression techniques are mostly used. Clustering focuses on forming groups with similar properties whereas in terms of classification predefined classes are used. A mathematical function is used to evaluate the data set when a regression algorithm is applied. The gained patterns are than used to describe groups. Members of the same group share a number of properties, or, in terms of behavioral advertising, interests.

5. Evaluation
During this stage the developed model will be evaluated. Tests have to be performed to check if the applied technique led to the desired result. It may not happen that a large amount of data is not categorized correctly. Thus,

[2] `http://www.crisp-dm.org/Process/index.htm` (accessed 05-05-11)

if there are still issues that have not been sufficiently treated, parameters of the data mining algorithm have to be changed or, in the worst case, another algorithm has to be chosen. If the results meet the expected criteria, a selection has to be made. Not all information is useful in a given situation. Custer [Cus04] states that an important phenomenon in this context is masking. Masking means that particular characteristics are correlated and it may be possible to use trivial characteristics as indicators of sensitive characteristics. For example, when people live in zip code area that has a high health risk, insurance companies may use the trivial information (zip code) as an indicator for the sensitive information (health status). If this is prohibited by law or not will not be discussed in this chapter, it is just an example to give a better impression of what is possible.

6. Deployment
During the last phase the repeatable data mining process has to be implemented and corresponding actions have to be defined. It can be described as a rule base with if-then statements. In terms of behavioral advertising, the different profile groups have to be linked with the corresponding advertisements that statistically match with the interests of the customers.

Finally, we have to mention that data mining with regard to behavioral advertising is an ongoing process. New group profiles will be created and existing ones will be updated or even deleted. It can occur that an individual will be categorized as a member of a certain group and one month later, because of additional information that is gathered, the same individual now belongs to another profile group.

2.2.2 Group Profiling

Group profiling can be done either by using data mining techniques or by doing empirical statistical research. Custer [Cus04] mentions that there may be differences when group profiling using data mining is applied. Data mining algorithms are much faster in generating group profiles than empirical statistical research. Also large amounts of data can be easily processed by data mining algorithms but there exists a risk of getting too many group profiles. Furthermore, data mining offers the possibility to investigate every possible relation. Thus, data mining algorithms generate many different hypothesizes. These hypothesizes are then checked against the data set which results in more different profile groups. Also masking which we discussed earlier can be applied by data mining algorithms. Custer mentions also some privacy concerns with regard to group profiling by data mining. Data mining techniques are less transparent and people who share some trivial information about themselves on the internet do not expect that someone can uncover sensitive information using a data mining technique.

Profile information types

Not every group profile consists of the same information structure. Borking et al. [BvAA98] describe five types of information.

1. Associations
2. Sequences
3. Classification
4. Clustering
5. Predictions

Associations are made when different characteristics are linked. For example, if-then structures with a certain probability are typical associations. Sequences and associations are the same except for the time issue. In case of sequences, time plays an important role. Some event occurs and later, for example two days, a follow-up event occurs. Classification deals with known groups to determine if certain characteristics can be used as an identifier for a group. Clustering takes as a starting point a characteristic, for example age. The goal is to identify different groups within the age category, like young people live in cities and middle-aged people live in the countryside. Thus, classification uses known groups as a starting point whereas clustering begins with known characteristics. Predictions are made for parameters which are not discovered yet. They are also similar to sequences but predictions focus on parameters and sequences on events.

Distributivity

We have to distinguish between distributive and non-distributive group profiles. According to Hildebrandt *"a distributive profile identifies a group of which all members share all the attributes of the group's profile"* [Hil08]. She mentions as an example the category of bachelors that all share the attribute of not being married. On the opposite side, a non-distributive profile consists of members that do not share all the attributes of the group's profile. We can assume that most group profiles that are used for behavioral advertising are non-distributive and, thus, not all members share all attributes. It is for a marketeer not relevant that some group profiles to not match the characteristics of an individual.

This can have a big impact on the treatment of an individual. In section 2.2.1, an insurance example is given. The used group profile assumes that all people living in a certain zip code area have a high health risk. This does not have to be the case because a high health risk depends also on, for example, age, gender or medical history. It is possible that in this area people will receive medical advertisements while they are young and totally healthy. This can also have substantial consequences when, for example, someone is looking for a life insurance.

Logic and Inference
Logic is used to determine whether an argument is valid or not. An argument is divided into a premise and a conclusion. The premise can exist of more than one statement. The examples are based on information from [Cus04]:

1. All information science students have to follow the course business rules. (premise)
2. Jan is an information science student. (premise)
3. Thus, Jan has to follow the course business rules. (conclusion)

The given example is an deductive argument. With regard to group profiling, deduction is often not possible because of limited information. The idea of deduction is to work from general statements to specific examples but in group profiling it is mostly the other way around. Specific examples are used to make predictions for a much greater group. Therefore induction is used more often because it leaves some space. An inductive example is the following:

1. These people are information science students. (premise)
2. These information science students have to follow the course business rules. (premise)
3. All information science students have to follow the course business rules. (conclusion)

This way of reasoning is not conclusive and thus not deductive because in case of deduction the conclusion must be true. In this example, it is still possible that there are information science students who do not have to follow the course business rules. The conclusion is only reliable if a lot of cases are examined.

Also a third method, abduction, exists:

1. All information science students follow the course business rules. (premise, hypothesis)
2. These students follow the course business rules. (premise)
3. These students are information science students. (conclusion)

Thus, induction reasons from specific to general and abduction reasons from result to cause. Abduction is less useful for group profiling because it suggests what may be the case. This will not say that advertising networks do not use abduction for profiling because suggesting what may be the case with a high probability is just what they want. Group profiles based on induction are in general less reliable than group profiles that are based on deduction because a lot of data is needed to proof that the inductive method is reliable. Induction can still be used to test how many members fully apply to the attributes of a group profile.

Finally, it has to be clear that from the moment a group profile exists, individuals who belong to that profile group will be treated different. Thus, after data mining and profile generation, group profiles will be applied to individuals. An individual does not have to share all of the characteristics of a particular group profile. This can influence the treatment of an individual in a positive but also in a negative way.

2.3 Stakeholders

This section deals with 3 different types of stakeholders: advertising networks, publishers and Internet Service Providers. Furthermore, we discuss what could happen if an Internet Service Provider starts his own advertising network. The main stakeholders, the user or data subject and the advertiser are already discussed in section 2.1.

2.3.1 Advertising Networks

This section deals with advertising networks. First, we will give a definition of the term advertising network and a general introduction. The second part deals with three advertising networks: Google, Yahoo! and Microsoft.

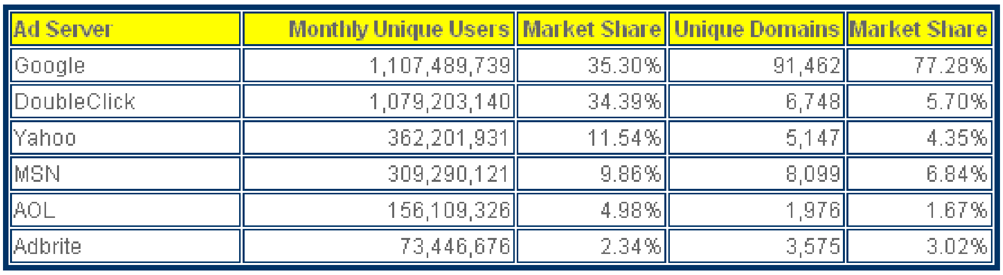
Many similar definitions of the term advertising network exist on the internet. We will handle the following one:

> *"Advertising network refers to an infomediary, which serves between a group (network) of web sites (which want to host advertisements) and advertisers which want to run advertisements on those sites. DoubleClick is the most well known advertising network. It can serve advertisements on any of its member sites, and can mine data from users who interact with those sites in order to serve more appropriate advertisements. An advertiser can buy a run of network package, or a run of category package within the network. The advertising network serves advertisements from its ad server, which responds to a site once a page*

> *is called. A snippet of code is called from the ad server, that represents the advertising banner."*[3]

Advertising networks analyze the behavior of internet users to enhance the value of their advertisements. Different revenue models such as cost per thousand (CPT), cost per click (CPC) or cost per action (CPA) exist [HN00]. CPT means per thousand loads of an advertisement, CPC means payment for each click on the advertisement and CPA means payment for each time a user clicks on the advertisement and makes a purchase or registers for a service.

The following companies are the key players with regard to market share:



Ad Server	Monthly Unique Users	Market Share	Unique Domains	Market Share
Google	1,107,489,739	35.30%	91,462	77.28%
DoubleClick	1,079,203,140	34.39%	6,748	5.70%
Yahoo	362,201,931	11.54%	5,147	4.35%
MSN	309,290,121	9.86%	8,099	6.84%
AOL	156,109,326	4.98%	1,976	1.67%
Adbrite	73,446,676	2.34%	3,575	3.02%

Figure 2.3: Ad Server Market Share[4]

The results were gathered in 2008 and are based on data from 68 million domains. Google has together with DoubleClick a market share of almost 70% when we consider the monthly unique users. It is even higher, around 84% when we consider the unique domains that are managed by Google and DoubleClick. Thus, it is clear that Google dominates the online advertising market. A New York Times reporter, Louise Story, wrote an article *Where Every Ad Knows Your Name* in 2008[5]. She asked four different advertising companies, Google, Microsoft, Yahoo! and AOL, if they could add her name to an advertisement by searching for her real name in their databases. Microsoft stated that they only could use her first name, AOL and Yahoo! could use her full name but only on their own sites and Google is not sure but they probably could, but most of the names of Google's customers are unknown. It is still a mystery what these companies are really capable of. The given answers represent what is stated in their privacy policies.

[3] `http://www.learnthat.com/define/view.asp?id=266`(accessed 11-05-11)

[4] `http://attributor.com/blog/get-your-fair-share-of-the-ad-network-pie/` (accessed 11-05-11)

[5] `http://bits.blogs.nytimes.com/2008/03/10/where-every-ad-knows-your-name/`(accessed 11-05-11)

Google

Google has developed an ad serving application AdSense for website owners. It gives them the opportunity to earn money by adding advertisements provided by Google to their websites. AdSense is very popular because money can be easily earned [Con08]. The advertisements are context sensitive based on the content of the site. The advertisements can be of different kind such as video, mobile content or RSS feeds. Sites that include AdSense advertisements provide the IP address of each user to Google and allow cookies for tracking purposes.

According to Greg Conti [Con08] AdWords is a fundamental part of Google‘s business model. AdWords is the main source of revenue. It provides advertisements each time a user uses Google‘s search engine. Thus, Google uses the search term to select appropriate advertisements. Companies use AdWords to create their own advertisements and to define keywords for which their ad will appear. The advertisement will also be displayed on web pages that earn money with AdSense.

DoubleClick is a subsidiary of Google and counts a large number of Fortune 500 companies as clients [Con08]. By acquiring DoubleClick, Google now displays not only search and simple text advertisements but also banner and video advertisements.

Yahoo!

Yahoo! has its own subsidiary Yahoo! Advertising. They claim on their homepage that they can reach the right customers at the right time using behavioral targeting[6]. Furthermore they give a short explanation about how Yahoo! BT works:

> *"Yahoo! BT goes beyond common rules-based segmentation or grouping of consumers by the sites they've visited. The tool is powered by sophisticated modeling technology based on extensive online interactions that include searches, page views, and ad interactions. With these models, Yahoo! identifies what consumers are interested in and predicts where they are in the buying process, thereby determining which consumers may respond best to your ad placements."* [7]

It is clear that they want to convince new customers that Yahoo! offers what

[6] http://advertising.yahoo.com/products-solutions/behavioral-targeting.html(accessed 12-05-11)

[7] http://advertising.yahoo.com/products-solutions/behavioral-targeting.html(accessed 12-05-11)

everyone needs. They do not mention privacy related issues and how they track people because this could create a negative connotation.

Also research is done by two Yahoo! employees on managing reach and frequency to maximize response rates in online advertising [NF09]. One of their findings is that up to 80% of ad clicks come from users who see a campaign 3 times or less.

Microsoft

Microsoft‘s advertising network uses Hotmail account information and the search history of Microsoft Live Search for profiling purposes. Thus, Microsoft uses its Hotmail service for profiling whereas advertising networks such as DoubleClick use cookie information from lots of other websites. Aaron Patrick, a Wall Street Journal reporter, published in 2006 the article *Microsoft ad push is all about you*:

> *"Here's how it works: If someone types in "compare car prices" on Live Search, Microsoft's computers note that the person is probably considering buying a vehicle. The computers then check with the list of Hotmail accounts to see if they have any information on the person. If they do, and an auto maker has paid Microsoft to target this type of person, the computer will automatically send a car ad when she next looks at a Microsoft Web page. As a result, people should see more ads that are of interest to them. "We know what Web sites they have visited and what key words they used," says Mr. Dobson. "We can deduce what their interests are." Microsoft says that in testing in the U.S., behavioral targeting increased clicks on ads by as much as 76 percent."* [8]

It would be interesting to examine if Hotmail users are aware of the fact that Microsoft uses their personal information for behavioral targeting.

All three advertising companies use search logs, website views and known information such as account details for advertising purposes. Thus, they make extensive use of behavioral targeting which raises a lot of privacy related concerns.

2.3.2 Publishers

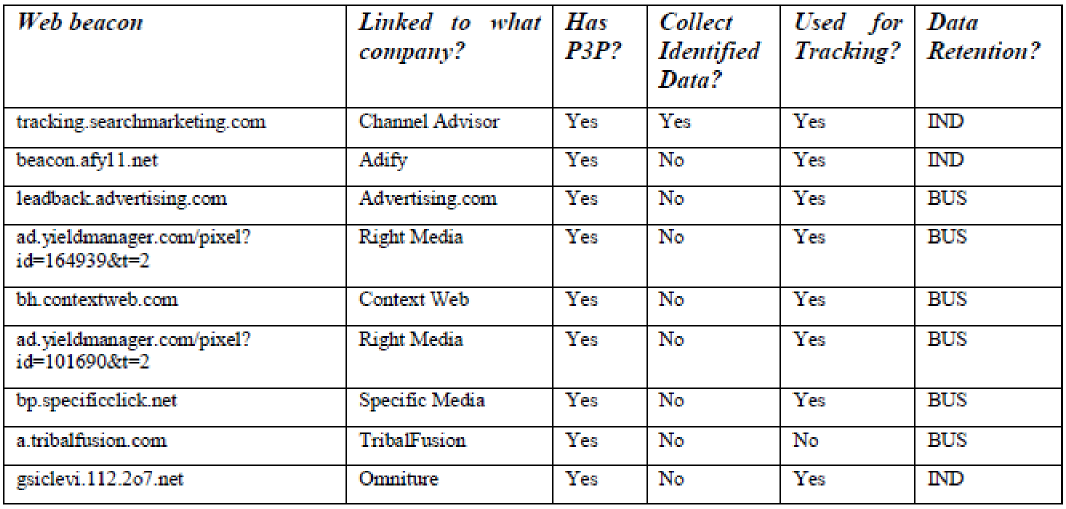
Nowadays, commercial websites and others such as blogs or communities earn money by adding third-party content to their sites. Users who visit such

[8] `http://www.post-gazette.com/pg/06361/749452-96.stm` accessed 16-05-11)

sites are often not aware of how many third-party companies get information about the visitor. We will give two examples, the first example is a case study done by Catherine Dwyer [Dwy09] who examined Levis.com and the second is an analysis of the msnbc.com website described in [Con08].

Levis.com

Dwyer [Dwy09] first examines the privacy policy of Levis.com. It describes how personal information will be treated and that this information will not be shared without the customer‘s consent. The collection of non-personal information is described in a separate category. They state that they collect customer traffic patterns, site usage and other information. Furthermore two third-party companies are mentioned in the privacy policy, one that collects anonymous information about website visits for advertising purposes and one that measures the use of the site.



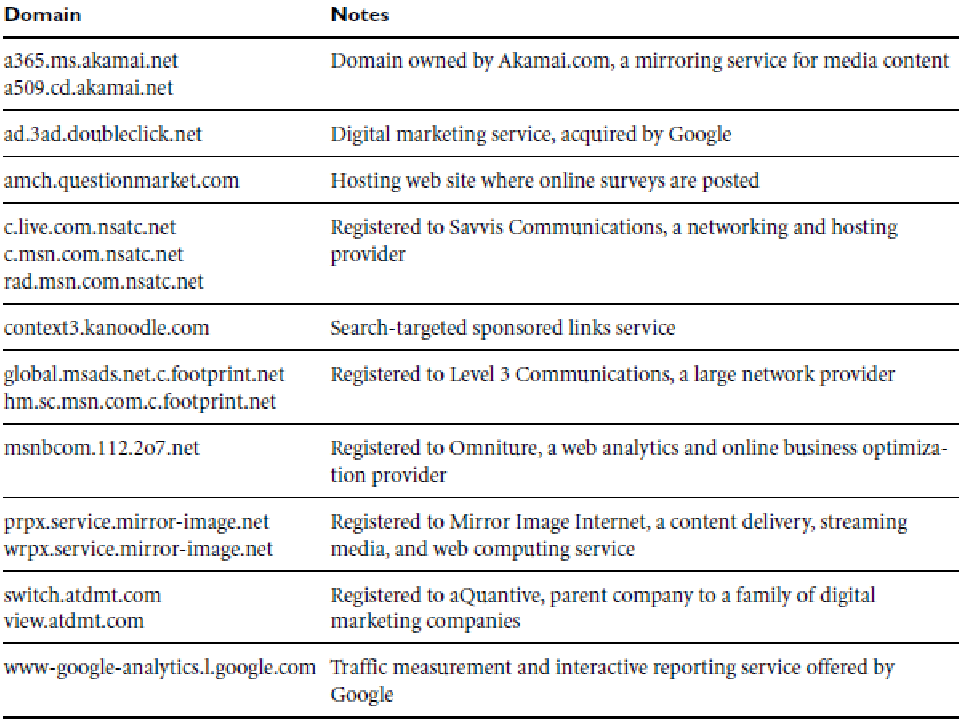
Web beacon	*Linked to what company?*	*Has P3P?*	*Collect Identified Data?*	*Used for Tracking?*	*Data Retention?*
tracking.searchmarketing.com	Channel Advisor	Yes	Yes	Yes	IND
beacon.afy11.net	Adify	Yes	No	Yes	IND
leadback.advertising.com	Advertising.com	Yes	No	Yes	BUS
ad.yieldmanager.com/pixel? id=164939&t=2	Right Media	Yes	No	Yes	BUS
bh.contextweb.com	Context Web	Yes	No	Yes	BUS
ad.yieldmanager.com/pixel? id=101690&t=2	Right Media	Yes	No	Yes	BUS
bp.specificclick.net	Specific Media	Yes	No	Yes	BUS
a.tribalfusion.com	TribalFusion	Yes	No	No	BUS
gsiclevi.112.2o7.net	Omniture	Yes	No	Yes	IND

Figure 2.4: Levis.com: Web beacons [Dwy09]

Dwyer used a Firefox plug in to log the content of the Levis homepage. One of her findings is that Levis homepage contains nine different web beacons (see section 2.4.2 for further explanation) that link to eight advertising companies. Eight beacons are used to track customers and one even collects identified data. None of the advertising companies are mentioned in the earlier discussed privacy policy. All beacons have a P3P policy. P3P is a protocol that can be implemented into websites to declare the purpose of data collection (see section 4.2.4). Three beacons will, according to their P3P, retain data for an indeterminate (IND) period of time. The remaining six collect data for business practices (BUS). According to the P3P specification companies who collect data for business practices have to mention that in the provider‘s privacy policy, Levis does not do that. This research examines just one website but we can assume that adding tracking content to commercial websites is a common practice.

Msnbc.com

The msnbc.com website which is examined in [Con08] contains 16 additional third-party domains from 10 different companies. Figure 2.5 is a table of third-party domains on msnbc.com.



Domain	Notes
a365.ms.akamai.net a509.cd.akamai.net	Domain owned by Akamai.com, a mirroring service for media content
ad.3ad.doubleclick.net	Digital marketing service, acquired by Google
amch.questionmarket.com	Hosting web site where online surveys are posted
c.live.com.nsatc.net c.msn.com.nsatc.net rad.msn.com.nsatc.net	Registered to Savvis Communications, a networking and hosting provider
context3.kanoodle.com	Search-targeted sponsored links service
global.msads.net.c.footprint.net hm.sc.msn.com.c.footprint.net	Registered to Level 3 Communications, a large network provider
msnbcom.112.2o7.net	Registered to Omniture, a web analytics and online business optimization provider
prpx.service.mirror-image.net wrpx.service.mirror-image.net	Registered to Mirror Image Internet, a content delivery, streaming media, and web computing service
switch.atdmt.com view.atdmt.com	Registered to aQuantive, parent company to a family of digital marketing companies
www-google-analytics.l.google.com	Traffic measurement and interactive reporting service offered by Google

Figure 2.5: Third-party sites visited when browsing msnbc.com [Con08]

The extent to which users get tracked by visiting one single website can be very high. Visiting msnbc.com creates 16 log files. It seems to be a common practice to track visitors of a website and spread this information to a large number of third-party companies. Most of the visitors do not know these third-party companies. It is a common practice that these companies operate in the background.

2.3.3 Internet Service Providers (ISPs)

ISPs differ, with regard to behavioral advertising, in one significant way from advertising networks. They have far more information about an individual.

All traffic is routed by ISPs and, thus, can also be examined by them. KPN, a Dutch telecommunication company, recently announced that it is the first operator in the world that uses deep packet inspection to precisely measure what mobile internet users do[9]. All information that is sent through the internet is divided into smaller packets. By using deep packet inspection (DPI) ISPs have full access to the packets and their content. This raises numerous privacy concerns because DPI gives ISPs the possibility to create very rich profiles containing personal information. Thus, ISPs do not have to make assumptions any longer. They can analyze every piece of information.

2.4 Tracking Methods

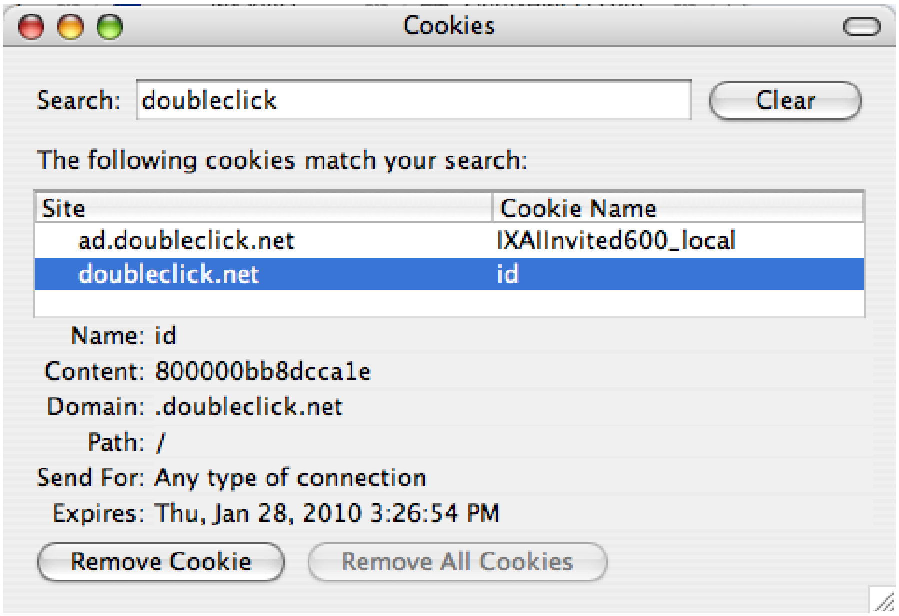
Multiple methods exist for tracking the online behavior of an individual. In the following, we will describe the most popular tracking methods that are used for behavioral advertising. The IP address of a user is less used to identify a user because it can change over time. Most advertising companies use (flash) cookies and web beacons to identify a user.

2.4.1 Cookies

Greg Conti says in his book that *"cookies are like the tracking darts scientists shoot into wild animals on nature documentaries"* [Con08]. This is a very good example because cookies do exactly the same in the context of behavioral advertising. Different cookie types exist but two types are used most of the time: persistent cookies and session cookies. Session cookies are not used for tracking because they only exist for a short period of time. A typical example for a session cookie is when someone does internet banking. Every time when the user logs in to his/her account, a session cookie is used. Session cookies make the interaction easier because otherwise, without a cookie, user had to log in again every time when the user makes a new request. Persistent cookies exist, as the name already states, for a long period of time. Thus, they are mostly used for tracking users because every time a user visits the website that issued the cookie, it will be logged on the web server. The same holds for cookies that are issued by advertising networks except that the user has no intention to visit the website of the advertising network. That kind of cookies are called third-party cookies and are issued because websites often contain embedded content of advertising companies [Con08]. We will describe the structure of a cookie using the following example cookie (figure 2.6) which is issued by DoubleClick.

All cookies have the same structure, a name field, a content field, a domain

[9] `http://webwereld.nl/nieuws/106656/kpn-luistert-abonnees-af-met-deep-packet-inspection.html` (accessed 23-05-11)



Search: doubleclick

The following cookies match your search:

Site	Cookie Name
ad.doubleclick.net	IXAllnvited600_local
doubleclick.net	id

Name: id
Content: 800000bb8dcca1e
Domain: .doubleclick.net
Path: /
Send For: Any type of connection
Expires: Thu, Jan 28, 2010 3:26:54 PM

Figure 2.6: Example Cookie which is issued by the advertising company DoubleClick (Google)[10]

field, a path field, a send for field and an expiration field. In case of behavioral advertising, the content field is the most important one because it includes an identifier. The life time of a cookie is determined by the company that issues the cookie. Greg Conti mentions in his book [Con08] that, for example, Google cookies expire after 2 years and Yahoo! cookies expire after 29 years but Google cookies are auto-extended with each visit to Google. We can assume that internet users visit Google far more than one time in two years and, thus, the cookie will never be discarded by the browser.

Finally, advertising companies are able to observe the behavior of an individual even when the individual visits multiple sites because of embedded content and the corresponding cookie which is provided by the advertiser. Fig 2.7 illustrates cross-site tracking by an advertiser.

As shown in figure 2.7, 6 visits of different websites create 6 log entries on the server of the advertising network because all directly visited websites include embedded content of the advertising network. Also HTTP referrer data gives the advertiser information about from which site a request originates.

[10] `http://www.mikeonads.com/2007/02/27/whats-really-in-my-cookie-cache/` (accessed 19-05-11)

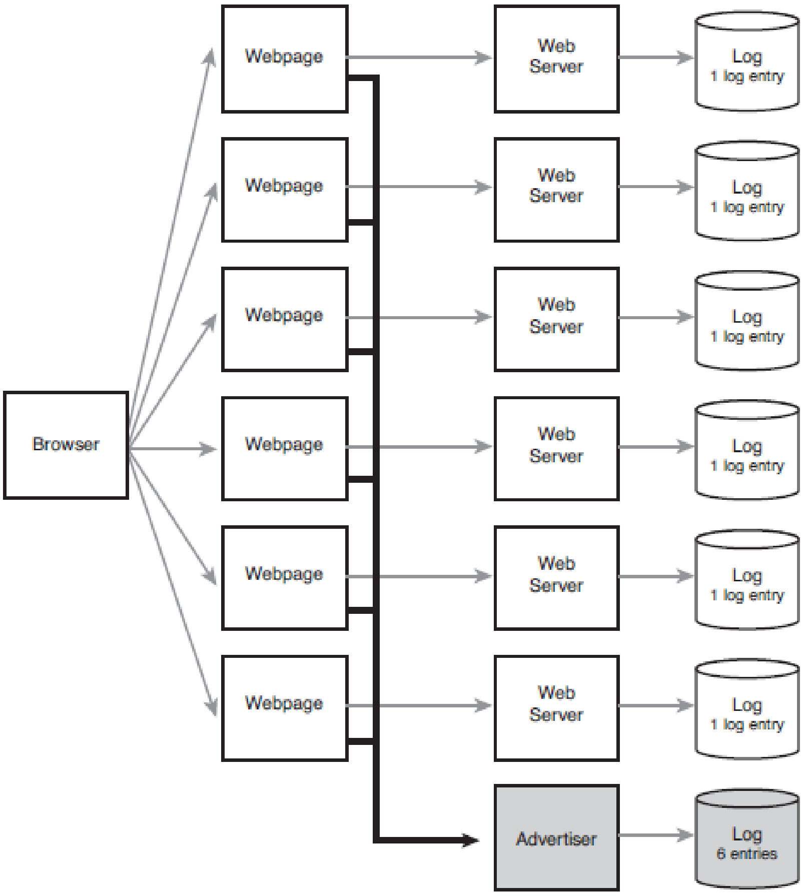


Figure 2.7: Cross-site tracking by an advertiser [Con08]

Adobe Flash Cookies
The Adobe Flash plugin is often used for animation and graphics. Adobe Flash uses a local data store which is called shared data objects by Adobe but it is also often called a Flash cookie. Adobe offers to their customers an online tutorial on how to implement flash cookies and track users[11].

2.4.2 Web Beacons (JavaScript)

According to Colin Bennett [Ben01] a web bug or web beacon is a graphic on a web page or in an email message that is designed to monitor who is reading the web page or email message. Web beacons are often invisible because they are graphics of the size 1 x 1 pixel with no color. They are usually written in JavaScript.Web beacons are stored in the cache of a browser which was

[11] `http://www.adobe.com/resources/richmedia/tracking/`(accessed 23-05-11)

originally designed to improve page loading speeds. The http header of a web beacon contains data fields with information that is used to facilitate tracking [Dwy09]. Web beacons are a typical form of page tagging. Using JavaScript offers the possibility to send extra information, such as the screen size of a user, together with the request for the web beacon to the server.

The example code[12] of a web beacon that can be added to an email message is given below:

```
<!DOCTYPE HTML PUBLIC "-//W3C//DTD HTML 4.0 Transitional//EN">
<HTML>
<HEAD></HEAD>
<BODY>
<IMG height=1 src="http://www.adoko.com/mybug.gif" width=1 target="_blank">
This email tests the use of a Web Bug.
</BODY>
</HTML>
```

The included 1 x 1 image is loaded when the user reads the email message. If someone wants to track multiple users using a web beacon, just adding a unique identifier, for example in the name field, would be sufficient. When we take a look at the log file of the web server we see the following entry:

```
2003-06-02 09:39:13
W3SVC110 NTXPW04 212.227.124.8
GET /mybug.gif - 80 -
212.24.161.236 HTTP/1.0 Mozilla/4.0+(compatible;+MSIE+6.0;+Windows+NT+5.1)
```

The user who reads the email has loaded the image at 09:39:13. Also the IP address and the browser type are logged.

2.4.3 Facebook Like Button

This section deals with behavioral advertising in social network communities. We examine the social network Facebook. Facebook can also be seen as an advertising network but there is one crucial difference between Facebook and other advertising networks: they have two sources of information, user profile information and tracking information.

Arnold Roosendaal [Roo] did research on the functionality of the ”Like” button which is introduced by the social network Facebook. Website owners can add the Like button to their websites. If a Facebook user clicks on the button a preview of the site will be send to his/her Facebook account. The button is for website owners a useful commercial tool because every time a

[12] http://www.adoko.com/webbugs.html (accessed 23-05-11)

Facebook user clicks on the button, the website will be posted on the wall of that user and thus visible to friends. Roosendaal examined the technical specifications of the Facebook Like button and he found out that adding the button to a website has a huge impact on privacy. Three different scenarios where examined:

1. A web user has a Facebook account.
2. A web user does not have a Facebook account.
3. A web user becomes a Facebook member.

A web user has a Facebook account

A unique user ID is assigned to each new Facebook account. This ID is stored in a cookie on the user's computer. The Facebook servers can recognize every user by processing the unique user ID. It does not make any difference when the same user logs in to Facebook from different devices. Thus, Facebook knows the identity of the visitor. This cookie is not only used when members want to login but it is also send to the Facebook servers when a website contains Facebook content such as the Like button. Every time when a user visits a website containing a Facebook Like button the cookie is sent to the server. Facebook is in a position to not only track users but also to identify users because of the unique user ID. In case of behavioral advertising, two different sources of information can be used and combined: account information such as interests, gender or age and tracking information.

A web user does not have a Facebook account

When a visitor of a website which contains Facebook content does not have an account, no cookie and user ID exists. It seems that the Facebook Like button does not issue cookies to users. The tool Facebook Connect can be used by website owners to make a connection between a Facebook account and a third-party website[13]. It is not necessary that this connection is made before a cookie can be assigned. Websites which have Facebook Connect automatically assign a Facebook cookie to every user that visits the site even when the user does not have a Facebook account. Every time when a website with Facebook content is visited, the cookie will be sent to the Facebook servers and, thus, tracking of non Facebook members is possible.

[13] http://developers.facebook.com/blog/post/108/ (accessed 28-05-2011)

A web user becomes a Facebook member

At the moment when someone joins the Facebook community the cookie which is assigned by Facebook Connect, is sent to the Facebook server. When the new member clicks on the button to create a new account, Facebook creates a new user ID and sends a new cookie containing that ID. We can assume that the earlier gathered data will be connected with the new user ID.

Finally, we can state that Facebook is not only able to track users but they can also create very rich personal profiles by combining tracking information with account information. Most of the Facebook members enter their real name when they join Facebook. Here, one important questions is: does Facebook process personal data when they combine these data sources for delivering targeted advertisements to their users?

2.4.4 Browser Fingerprint

A browser fingerprint, also called device fingerprint, consists of information about the hardware and software of a computer. Also configuration details such as resolution are included. The Electronic Frontier Foundation (EFF), a digital rights advocacy, started a project Panopticlick[14] to test if a browser, and thus most likely a single user, has a unique digital fingerprint. Peter Eckersley from EFF mentions in his paper [Eck10] that even when browser fingerprints change, because of configuration changes, it is still possible to track and identify them because of the amount of information that they reveal. He also states that browser fingerprints have the potential to identify a user and, thus, should be seen as personal identifiable information.

2.5 Location-Based Data

This section deals with location-based services (LBS). We will briefly describe what location-based services are, how LBS can be related to our research topic and what kind of privacy concerns play a role. We will not discuss LBS in detail because this would be outside the scope of this thesis.

LBS offer information, entertainment or social media on a mobile phone by using the geographical position. Many different LBS exist, for example, for navigation, near gas stations or restaurants in the neighborhood [SNE]. These services are embedded in applications that are developed for mobile phones or other mobile devices. Three tracking methods exist: base station data, GPS and WIFI [Parb]. Mobile devices are always connected to a base station when they are switched on. Data that is gathered via a base station

[14] `https://panopticlick.eff.org/` (accessed 23-08-2011)

is called base station data. The second tracking method is via GPS and the third one via a WIFI network that has a connection with the mobile device. WIFI access points are similar to base stations. Access points have unique IDs and services exist that deliver the corresponding location for each ID [Parb].

On the one side, LBS offer huge opportunities to people who use them. People can easily discover if there are any events coming up in their neighborhood or check which gas station is the cheapest. On the other side, every time when someone uses LBS, data is transferred about the geographical position and, maybe, also personal data. We can assume that traveling profiles can be easily generated by analyzing the geographical positions. Furthermore, preferences and the action history can also be used [SNE]. Steiniger [SNE] mentions that such analysis helps to get a perfect customer model but it can also raise user fears. Thus, users should be informed about the information that is collected. Dobson and Fisher [DF03] argue that this kind of tracking devices make real-time control possible. Therefore, they claim that governments and industries must develop LBS safeguards. The Article 29 Data Protection Working Party (see section 3.5) recently discussed in their *Opinion 13/2011 on Geolocation services on smart mobile devices* [Parb] privacy risks and how to deal with LBS. The Working Party recognizes that smart mobile devices are very intimately linked to a specific individual and that most individuals do not share them with others. Emails, SMS, private pictures and the contact list are stored on them. Tracking can be done for all kind of purposes such as behavioral advertising or monitoring of children. According to the Working Party, different obligations for the different stakeholders (developers, applications providers, social networks etc.) exist. The Data Protection Directive (see section 3.3) is applicable under all circumstances. The ePrivacy Directive (see section 3.4) is only applicable when base station data is processed by telecom operators. Furthermore the Working Party recommends that consent must be specific and cannot be obtained by general terms and conditions. Finally, they suggest an opt-in policy. Location services must be switched off by default.

Chapter 3

Data Protection & Privacy

This chapter deals with European law that provides a legal ground for behavioral advertising and, in particular, the use of cookies and future tracking technologies that allow the processing of (personal) data. In the first two sections, we examine the European Convention on Human Rights [ECHR] and the Charter of Fundamental Rights of the European Union [CFREU]. Furthermore, we introduce two EU directives that deal with data protection and privacy. We will discuss the most relevant articles of both directives. Section 3.5 discusses the task and opinions of the Article 29 Data Protection Working Party and section 3.6 deals with data protection and privacy in the Netherlands. Finally, in the last section, we will give an overview of achievements and unsolved problems with regard to EU legislation.

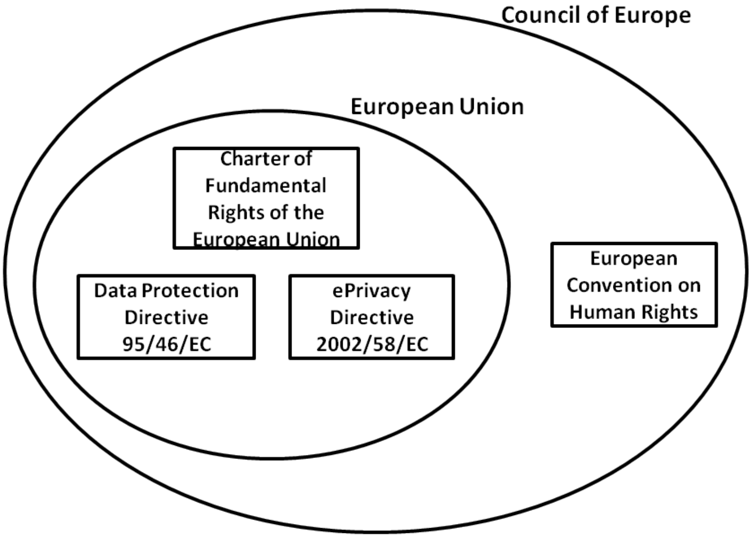

Council of Europe
European Union
Charter of Fundamental Rights of the European Union
Data Protection Directive 95/46/EC
ePrivacy Directive 2002/58/EC
European Convention on Human Rights

Figure 3.1: Applicable Law

The different laws and jurisdictions are illustrated in figure 3.1. The European Convention on Human Rights [ECHR] is at the top. Citizens of member states of the Council of Europe can go to court when they feel that their rights have been violated after having exhausted national remedies. Member States of the European Union must also comply with the Charter of Fundamental Rights of the European Union and have to implement the Data Protection Directive 95/46/EC and the ePrivacy Directive 2002/58/EC in national law.

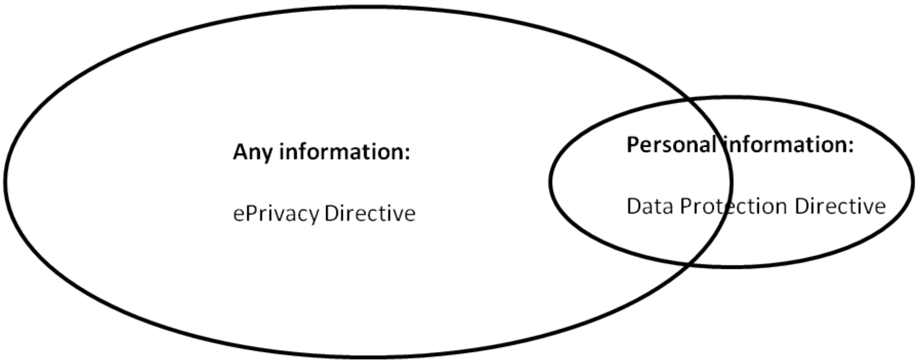


Figure 3.2: Relation between both directives

The ePrivacy Directive 2002/58/EC is always applicable in case of public electronic communication. This is illustrated in figure 3.2. The Data Protection Directive is only applicable when personal data is being processed.

3.1 European Convention on Human Rights

The European Convention on Human Rights [ECHR] deals with privacy in Article 8:

> **Article 8 - Right to respect for private and family life**
> 1. Everyone has the right to respect for his private and family life, his home and his correspondence.
> 2. There shall be no interference by a public authority with the exercise of this right except such as is in accordance with the law and is necessary in a democratic society in the interests of national security, public safety or the economic well-being of the country, for the prevention of disorder or crime, for the protection of health or morals, or for the protection of the rights and freedoms of others.

Article 8 is divided into two parts. In the first part, the right to privacy is defined. The second part deals with exceptions that can overrule the right to privacy. It is in principle only applicable to relations between governments and citizens.

Sottiaux [Sot08] mentions that Article 8(1) protects four different interests that are closely connected. According to Sottiaux, *"the Convention organs have explicitly declined to give an exhaustive definition of private life, but have instead identified different aspects of human life falling withing the ambit of this notion"* [Sot08] He analysis Article 8 by examining court cases that deal with it. In Article 8, four requirements have to be satisfied before it is legal to violate Article 8:

1. A violation must have a basis in law.
2. A violation must be necessary in a democratic society.
3. The measures that violate must have a legitimate aim.
4. The violation must be proportional in relation to the legitimate aim.

It must be clear and accessible, and safeguards must be provided. Furthermore, there must be pressing social need and a legitimate aim (Article 8(2)).

It is, for example, allowed to take pictures during a public demonstration. The private life is not fully protected from publicity. Article 8(2) deals with limitations. Only when it is *in accordance with law* and *necessary in a democratic society* interference with privacy interests are allowed [Sot08]. Third parties that possess private information of citizens fall within the scope of Article 8. It is not possible to go to court and claim that advertising networks are violating Article 8 because, as we already stated, it is only applicable to relations between governments and citizens. However, citizens can claim that the government does not provide privacy protecting measurements that sufficiently deal with behavioral advertising (with regard to Article 8). Thus, Article 8 offers a broad framework and also applies to behavioral advertising.

Behavioral advertising takes place between private parties (organizations) and citizens. We discuss in the next sections two EU directives that deal with how national legislation can provide informational privacy and data protection.

3.2 Charter of Fundamental Rights of the European Union

EU member states have to fully comply with the Charter of the Fundamental Rights of the European Union [CFREU]. National law has to be consistent with what is stated in the charter. Article 7 and 8 deal with privacy and personal data:

> **Article 7 - Respect for private and family life**
> Everyone has the right to respect for his or her private and family life, home and communications.

Article 7 is very similar to Article 8(1) of the European Convention on Human Rights but there is one crucial difference. In Article 7, the term communications is used instead of correspondence. The term correspondence is more related to letters whereas the term communication incorporates all kinds of information exchanges.

> **Article 8 - Protection of personal data**
> 1. Everyone has the right to the protection of personal data concerning him or her.
> 2. Such data must be processed fairly for specified purposes and on the basis of the consent of the person concerned or some other legitimate basis laid down by law. Everyone has the right of access to data which has been collected concerning him or her, and the right to have it rectified.
> 3. Compliance with these rules shall be subject to control by an independent authority.

Article 8 deals with the protection of personal data. The content of Article 8 is more concrete but it still leaves space for different interpretations. When we take a closer look at what is stated in the Data Protection Directive 95/46/EC and the ePrivacy Directive 2002/58/EC, we see that the content of Article 8 forms a unique basic right for both directives.

3.3 Data Protection Directive (95/46/EC)

The European Union enacted in 1995 the *Directive 95/46/EC on the protection of individuals with regard to the processing of personal data and on the free movement of such data.* In section 3.3.1 we will give a general introduction to the directive. Section 3.3.2 deals with implications for behavioral advertising.

3.3.1 General Introduction

Scope

In Article 2 of the directive definitions of the corresponding terms are given. The following definitions are most important in case of behavioral advertising:

> *(a) 'personal data' shall mean any information relating to an identified or identifiable natural person ('data subject'); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity;*
> *(b) 'processing of personal data' ('processing') shall mean any operation or set of operations which is performed upon personal data, whether or not by automatic means, such as collection, recording, organization, storage, adaptation or alteration, retrieval, consultation, use, disclosure by transmission, dissemination or otherwise making available, alignment or combination, blocking, erasure or destruction;*
> *[...]*
> *(d) 'controller' shall mean the natural or legal person, public authority, agency or any other body which alone or jointly with others determines the purposes and means of the processing of personal data; where the purposes and means of processing are determined by national or Community laws or regulations, the controller or the specific criteria for his nomination may be designated by national or Community law;*
> *(e) 'processor' shall mean a natural or legal person, public authority, agency or any other body which processes personal data on behalf of the controller;*
> *(f) 'third party' shall mean any natural or legal person, public authority, agency or any other body other than the data subject, the controller, the processor and the persons who, under the direct authority of the controller or the processor, are authorized to process the data;*
> *[...]*
> *(h) 'the data subject's consent' shall mean any freely given specific and informed indication of his wishes by which the data subject signifies his agreement to personal data relating to him being processed.*

A distinction is made between controller, processor and third party. Only the controller is defined as the entity that determines the purpose. The processor processes personal data on behalf of the controller and may not process the data for other purposes. Processing can also be done by third parties. Therefore, the term third party is introduced. A clear hierarchy exists with the controller at the top and third parties at the bottom. Further, the data subject's consent is defined. The data subject has to agree with the collection of personal data first before personal data may be collected and processed.

In Article 3 the scope of the directive is defined. Directive 95/46/EC *"shall apply to the processing of personal data wholly or partly by automatic means, [...]"*. We have to remark that the directive is only applicable in case of processing of personal data which is defined in Article 2(a).

Personal data may only be processed when the following is guaranteed: *proportionality*, *transparency* and a *legitimate purpose.*

Proportionality

Article 6 deals with proportionality and in Article 6(2) is stated that the controller has to ensure that personal data is (Article 6(1)):

(a) processed fairly and lawfully

(b) collected for specified, explicit and legitimate purposes and not further processed in a way incompatible with those purposes [...]

(c) adequate, relevant and not excessive [...]

(d) accurate, and where necessary, kept up to date [...]

(e) kept in a form which permits identification of data subjects for no longer than is necessary for the purposes for which the data were collected [...]

Also Article 8, 14 and 15 deal with proportionality. Article 8 prohibits *"the processing of personal data revealing racial or ethnic origin [...] health or sex life"* except when the data subject has given his explicit consent. Advertising networks that possess huge amounts of data can reveal such data by using data mining techniques. Thus, Article 8 does not directly deal with proportionality. In the example that we used in chapter 2 it is, thus, prohibited by law to process the data. Article 14 gives data subjects the right to object the processing of personal data for the purpose of direct marketing. Advertising networks that want to collect personal data, must not only inform the data subject about his rights (Article 10, 11 and 12) but they must also give the data subject the right to object. Article 15 states that a person has the right not to be subject to a decision which is based solely on automated processing except when the individual decision is taken in the course of entering into a contract. Group Profiling is based on automated processing and, thus, data subjects should have the right to object. This right is only be given in specific cases that form legal consequences or in the case of significant influence. Thus, it is more an exception than a permanent right. This is also mentioned by Bygrave who even goes one step further and says that what is promised in Article 15 is tarnished by the complexity and numerous ambiguities [Byg01]. Furthermore, a large number of conditions have to be met first before the right applies.

Transparency

The second principle, transparency, is incorporated in Article 7, 10, 11 and 12. Personal data processing must not only be in accordance with Article 6. Furthermore, one of the requirements of Article 7 have to be met. For example, consent has to be given first or the processing is necessary for the performance of a contract or for compliance with legal obligations etc. In practice, a gray area in law exists. Most data controllers are successful in arguing that their manner of data processing forms an exception and, thus, consent is not necessary. When they state that they do no process personal data, the whole directive does not apply.

Article 10, 11 and 12 deal with information that has to be given to the data subject and with the right of access. The identity of the controller and the purposes of processing must be provided to the data subject. The right of access includes that the data subject has the right to obtain from the controller information about *"the knowledge of the logic involved in any automatic processing of data concerning [. . .]"* (Article 12).

Legitimate Purpose

The third principle, legitimate purpose, is treated in Article 6(1)(b). Personal data may only be processed for a specific purpose and may not be processed for other purposes without consent of the data subject.

The Data Protection Directive also deals with the transfer of personal data to third countries. In Article 25(1) is defined that personal data may only be transfered to third countries that ensure *an adequate level of protection.* Furthermore, in Article 25(2) is mentioned that the adequacy shall be assessed in the light of all the circumstances surrounding a data transfer. Article 25 does not give a clear guideline on how to determine the level of protection that can be provided by a third country. Advertising network providers often operate in different countries and, thus, have to comply with Article 25 when they want to transfer personal data to a third country[1]

[1] This involves the policy of safe harbours and the complex solution for the transfer of data to the US.

3.3.2 Implications for Behavioral Advertising

The scope of Directive 95/46/EC is limited to processing of personal data. It is not applicable in case of anonymous data processing. Most advertising networks say that they do not collect personal information. In case that an advertising network collects personal information, data subjects have, according to Article 12, the right to ask that advertising network what kind of techniques they use when personal data is being processed. Automated individual decision making (Article 15) can affect the data subject in a way that is advantageous or disadvantageous compared with a decision, based on the same information, made by a person. Thus, automated decision making can produce positive and negative side effects. This has also some implications for behavioral advertising because data subjects could be excluded from advertisements, or the offered products have, for example, a different price for a certain group of data subjects. We have to remark that the lawmakers use the term *data subject* in all articles except for Article 15. In Article 15 the term *every person* is used. This change in terminology makes it even more difficult to understand the interdependence between all articles.

Schreurs et al. [SHKV08] mention that it may be important that the Data Protection Directive should apply to group profiling because it could hinder the construction of profiles. Furthermore, Schreurs et al. [SHKV08] give two more arguments for application of the Data Protection Law to group profiling. First, a group profile can be applied to an individual that was not involved during the construction of the group profile. Once used the applied profile may become personal data. Second, the Data Protection Directive is only applicable to group profiling where personal data is used. This makes it even more difficult to determine whether the Data Protection Directive is applicable.

3.4 ePrivacy Directive (2002/58/EC)

The European Parliament and the Council enacted in 2002 the ePrivacy Directive 2002/58/EC. This directive was amended by Directive 2009/136/EC in 2009. We will first give a general introduction and, in the second part of this section, we will discuss our findings with regard to behavioral advertising.

3.4.1 General introduction

The ePrivacy Directive focuses on data protection and privacy obligations for providers of public communication services. It deals not only with personal information, confidentiality and privacy but also with spam, web bugs and cookies. As discussed in chapter 2, cookies are mainly used to identify an individual in behavioral advertising. The ePrivacy Directive also applies to legal persons (Article 1(2)), whereas in the Data Protection Directive only individuals are treated.

The scope and aim of this directive is declared in Article 1(1). The directive has at its goal to protect the *"fundamental rights and freedoms, and in particular the right to privacy and confidentiality, with respect to the processing of personal data"* (Article 1). Both directives, the Data Protection Directive and the ePrivacy Directive, use the term *personal data* but relevant protections offered by the ePrivacy Directive exist that concern data and thus not necessarily personal data. Further, in Article 1(2) is stated that the ePrivacy Directive particularizes and complements the Data Protection Directive. Thus, a clear relation between both directives is made.

In Article 2 used terms are defined. We have to note that a user is defined as a *"natural person using a publicly available electronic communications service, for private or business purposes, without necessarily having subscribed to this service"*. Thus, all users are covered by this directive. Also the concerned services (Article 3) *"publicly available electronic communications services"* are mentioned. Security measurements must be provided for all services (Article 4) and confidentiality (Article 5) must be guaranteed. Cookies are covered in Article 5(3):

> *"3. Member States shall ensure that 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, in accordance with Directive 95/46/EC, inter alia, about the purposes of the processing. This shall not prevent any technical storage or access for the sole purpose of carrying out the transmission of a communication over an electronic communications network, or as strictly necessary in order for the provider of an information society service explicitly requested by the subscriber or user to provide the service."*

Thus, consent can only be given after clear and comprehensive information is provided to the user. Storing information in the terminal equipment of a user in itself is harmless but accessing the stored information raises lots of privacy concerns. Also recital 24 and 25 of Directive 2002/58/EC and recital 66 of Directive 2009/136/EC discuss the use of cookies. The three recitals are discussed in section 3.4.2.

Traffic data (Article 6) must be deleted or made anonymous after it is no longer needed. There is an exception in case a user has given his consent but the user must be informed about the purpose and period of time the data will be retained. The Data Retention Directive 2006/24/EC deals with data retention in detail. In most cases, data must be stored for not less than six months and not more than two years (Article 6). We will not discuss the Data Retention Directive in detail because this would fall outside the scope of this thesis.

Address data may only be used for commercial purposes after the user has given his consent (Article 13).

Summarized, we can say that the directive claims an opt-in solution. No data may be collected before consent is given but it does not give a clear guideline on how to achieve informed consent. Furthermore, the directive only deals with tracking methods such as cookies. We can assume that advertising networks will look for new tracking technologies when the obligations are getting too strict. For example, Deep Packet Inspection can be used for behavioral advertising purposes and is not mentioned in the ePrivacy Directive.

3.4.2 Implications for Behavioral advertising

Advertising networks use cookies and web bugs to identify users. Recital 24 of Directive 2002/58/EC deals with storing information in the terminal equipment of users:

> *"Terminal equipment of users of electronic communications networks and any information stored on such equipment are part of the private sphere of the users requiring protection under the European Convention for the Protection of Human Rights and Fundamental Freedoms. So-called spyware, web bugs, hidden identifiers and other similar devices can enter the user's terminal without their knowledge in order to gain access to information, to store hidden information or to trace the activities of the user and may seriously intrude upon the privacy of these users. The use of such devices should be allowed only for legitimate purposes, with the knowledge of the users concerned."*

According to recital 24, it is only allowed to store information in the user's terminal *with the knowledge of the users concerned.* We have to note that this recital refers to *any information stored on such equipment.* It seems that this directive does not distinguish between personal and anonymous information. Debussere [Deb05] argues that the term *any information* also applies to Article 5(3) and, thus, it seems that this directive applies to any information that is stored on a computer. Informed consent is a prerequisite before information may be collected or stored in the user's terminal. Also user tracking which is a fundamental part of behavioral advertising is treated by this recital and should only be allowed after consent is given.

Recital 25 explicitly addresses the use of cookies:

> *"However, such devices, for instance so-called 'cookies', can be a legitimate and useful tool, for example, in analysing the effectiveness of website design and advertising, and in verifying the identity of users engaged in on-line transactions. Where such devices, for instance cookies, are intended for a legitimate purpose, such as to facilitate the provision of information society services, their use should be allowed on condition that users are provided with clear and precise information in accordance with Directive 95/46/EC about the purposes of cookies or similar devices so as to ensure that users are made aware of information being placed in the terminal equipment they are using. Users should have the opportunity to refuse to have a cookie or similar device stored in their terminal equipment. [...] The methods for giving information, offering a right to refuse or requesting consent should be made as userfriendly as possible. Access to specific website content may still be made conditional on the well-informed acceptance of a cookie or similar device, if it is used for a legitimate purpose."*

Cookies are described as useful tools in analyzing website design and advertising. Thus, this recital also recognizes cookies in the context of behavioral advertising. Recital 25 also expresses the need for *clear and precise information* and refers to the Data Protection Directive. We have to mention that, according to recital 25, users should have the opportunity to refuse cookies and that the methods for requesting consent should be *as userfriendly as possible* and the should offer a right to refuse. Recital 25 is still meant to be an opt-out whereas the new Article 5(3) and 6(3) require opt-in. Also recital 66 is still meant to be an opt-out.

The ePrivacy Directive 2002/58/EC does not give any guidelines on how informed consent can be achieved. The amending Directive 2009/136/EC deals with achieving informed consent in recital 66:

> *"Third parties may wish to store information on the equipment of a user, or gain access to information already stored, for a number of purposes, ranging from the legitimate (such as certain types of cookies) to those involving unwarranted intrusion into the private sphere (such as spyware or viruses). It is therefore of paramount importance that users be provided with clear and comprehensive information when engaging in any activity which could result in such storage or gaining of access. The methods of providing information and offering the right to refuse should be as user-friendly as possible. [...] Where it is technically*

> *possible and effective, in accordance with the relevant provisions of Directive 95/46/EC, the user's consent to processing may be expressed by using the appropriate settings of a browser or other application. [...]"*

Recital 66 explicitly addresses third parties, such as advertising networks, who wish to store information on the equipment of a user. On the one side, clear and comprehensive information is of *paramount importance* but, on the other side, using the appropriate browser settings is sufficient for expressing consent. We can assume that most internet users choose the default settings when installing a browser. Thus, we can state that consent is more given by browser vendors and less by internet users because they decide how the default privacy and cookie settings look like.

Maybe, on the one side, it can be argued that consent is given by installing a browser und choosing the appropriate privacy settings but, on the other side, we cannot say that clear and comprehensive information is provided to the user because privacy settings are often explained in terms of low, medium or high privacy. Thus, both directives give indications on how to deal with the problem of achieving informed consent in behavioral advertising but they do not offer a clear guidance. Advertising networks can argue that some points need to be further discussed and choose an interpretation that fits best with their own wishes. The Article 29 Data Protection Working Party was set up under Article 29 of Directive 95/46/EC. We will discuss the task of the Data Protection Working Party in the following section.

3.5 Article 29 Data Protection Working Party

The Article 29 Data Protection Working Party is an independent European advisory body on data protection and privacy. Data protection authorities of the member states have an representative in the Article 29 Data Protection Working Party. Its tasks are to give advices to the EU member states and to promote that every member state implements the data protection and privacy directives in the same way. We will use the abbreviation *Working Party* when we refer to the Article 29 Data Protection Working Party. The working party itself is a legally instituted public law body, but its opinions do not have legal force. The Article 29 Working Party has been instituted in Article 29 of the Data Protection Directive 95/46/EC, for its precise competences we refer to Article 29 and 30 of the Directive.

3.5.1 Legal Framework

The Working Party adopted in June 2010 the *Opinion 2/2010 on online behavioral advertising* [Parc]. This opinion deals with concerns regarding the right to privacy and data protection in behavioral advertising. Furthermore, they provide a legal framework in which the roles and obligations of the different players are discussed.

According to the Working Party, Article 5(3) of the ePrivacy Directive is fully applicable in case of behavioral advertising because a tracking cookie is information that is stored in the data subject's terminal equipment and it is accessed by advertising network providers each time the data subject visits a partner website. Thus, advertising network providers have to comply with Article 5(3). They also distinguish between non-personal and personal data whereas, in the second case, Directive 95/46/EC will also apply. The methods that are used in behavioral advertising, allow advertising networks to collect IP addresses and unique identifiers (cookies). The Working Party argues that, even when the real name is unknown, it is still possible to identify a data subject because the identity can be 'singled out' [Parc]. Furthermore, it is possible to link the collected data with directly identifiable data, for example, when a data subject registers for a service. Hence, both directives apply in case of behavioral advertising.

Players

The Working Party also defines roles and responsibilities for three different players: advertising network providers, publishers (website owners) and advertisers (the company that offers a product or service).

Advertising network providers have always to comply with Article 5(3). According to the Working Party, it is not relevant if a data controller or data processor places or accesses a cookie. Informed consent should always be given. Special obligations exist when an advertising network provider processes personal data. In that case, the advertising network provider is a data controller and has to comply with Directive 95/46/EC.

Publishers rent a part of their website space to advertising network providers. Publishers themselves normally do not send tracking cookies and collect data. The Working Party notes that it is the browser of the data subject that sends information to the advertising network provider when the website of a publisher is visited. Publishers trigger the transfer of information (for example the IP address) and, thus, they have also responsibilities. The framework clearly says that publishers have to inform the data subject about the processing of data. Publishers are called joint controllers when they collect and transmit data to advertising network providers. The Working Party recommends that contracts between advertising network providers and publishers should be analyzed on a case by case basis. Thus, responsibilities can differ between publishers.

Advertisers are, according to the Working Party, independent data controllers when they observe a campaign, collect data, such as demographic data, and combine the collected data with registration data. The Working Party focuses on behavioral advertising carried out by advertising network providers and does not deal with obligations of advertisers.

Informed Prior Consent

The Working Party gives some recommendations concerning the term informed consent in Article 5(3) of amended ePrivacy Directive. According to what is stated in Article 5(3), consent must be obtained before someone places a cookie or information in the terminal equipment of a user. This can only be achieved if prior, clear and comprehensive information has been given to the user.

Browser settings can be configured that they reject third-party cookies which are used for behavioral advertising. Publishers and advertising network providers say in their privacy policy that third-party cookies are used for behavioral advertising and that users can reject them by using the appropriate browser settings. The Working Party explicitly states that this method does not meet the requirements of Article 5(3) because *prior* consent is not given and that recital 66 of the amended ePrivacy Directive is not an exception. Recital 66 only expresses that consent can be given in different ways. It seems to be difficult to determine when browser settings comply with the directives. According to the Working Party, this is difficult to realize because of the following:

1. Average data subjects have no expertise with regard to browser settings and they are mostly not aware of that someone tracks their online behavior. Three major browsers accept all cookies by default.

2. It should not be able to bypass an earlier made choice. Flash cookies can recover cookies that are deleted by the data subject.

3. Data subjects are often not aware of the purpose and use of cookies. Browsers that accept all cookies by default imply that users accept future processing.

Thus, clear and comprehensive information must be given before a cookie is being placed. Browsers may not accept any cookie before the data subject has given his consent. This is only sufficient in case that also information about the advertising network provider and the purpose is given. All criteria of Directive 95/46/EC have to be met.

Opt-out options are often offered to data subjects. This includes that advertising network providers give data subjects the opportunity to visit their website and to opt-out from behavioral advertising. The required prior informed consent is still not met because data is already gathered before a data subjects chooses to opt-out.

Prior opt-in consent is, according to the Working Party, the best solution to achieve what is stated in the directives. This can be done by showing the data subject specific messages and to ask for permission to set a cookie. The Working Party suggests that consent has to be renewed, for example, every year. Thus, cookies will have a limited lifespan and there must be a possibility to easily remove a cookie at any time.

Transparency

The Working Party gives some recommendations [Parc] on how to make behavioral advertising more transparent for data subjects. It is clear that the identity of the advertising network provider and purpose of processing should be provided to the data subject. This should be, according to recital 25 of the ePrivacy Directive, *as user friendly as possible.* The Working Party supports icons that are placed around advertisements by advertising network providers which give additional information. Data subjects have to be informed periodically about that they are monitored. The Working Party suggests that advertising network providers introduce a symbol that clearly indicates that monitoring takes place.

It is discussible who is responsible for providing information to the data subject. The Working Party recommends that both advertising network provider and publisher work together. It is not necessary that information is provided twice. It is more important that clear agreements are made about who is responsible and how the information is provided. Additional information about responsibilities and obligations of the data controller and data processor is given by the Working Party in *opinion 1/2010 on the concepts of "controller" and "processor"* [Para].

3.5.2 Personal Data

The Working Party published in 2007 the *Opinion 4/2007 on the concept of personal data* [Pard]. The Working Party recognizes that different interpretations among the member states exist. Uncertainty and diversity in practice exists. According to the Working Party, the European lawmaker adopted a broad notion of the term *personal data.* The right to privacy can be seen as a boundary that limits the notion. Therefore, data protection authorities play an important role because they have to apply what is stated in the directives. In opinion 4/2007, the term personal data is divided into four building blocks:

1. *any information*

2. *relating to*

3. *identified or identifiable*
4. *natural person*

The Working Party suggests a wide interpretation of the term *any information*. Thus, subjective and objective information about a person should be considered as personal data. Whether information relates to an individual or not must be determined by examining the content and purpose or result of the data processing task. The third element deals with conditions under which an individual is identifiable. The Working Party suggests that specific case analysis is needed because context and circumstances often play an important role. The last element deals not only with living persons but also with legal or deceased persons. The Directive is, thus, not only applicable to individuals that are natural persons but also to legal persons such as organizations.

3.5.3 Conclusion

The Working Party suggests that advertising network providers should work together with browser developers. Privacy by design could be a solution but does it really solve the entire problem? Opt-out solutions cannot guarantee that prior informed consent is given. Hence, opt-in solutions are the preferred ones. Also the purpose of data processing should be clearly defined in the information that is given to the data subject. Betsy Masiello and Alma Whitten, both working for Google, state in their article [MW10] that *unique opportunities to tackle the most difficult contemporary social problems* in the age of information abundance exist. According to them, problems can be solved by combining data of different sources using data mining techniques, but this raises a lot of privacy concerns. Therefore, it is necessary to protect data subjects by forcing data controllers to comply with the purpose limitation principle stated in Article 6(1)(b) of Directive 95/46/EC.

3.6 Data Protection and Privacy in the Netherlands

The Dutch Telecommunication Act (telecommunicatiewet) aims to protect the rights of citizens in the electronic communication sector. The Dutch government recently discussed an amendment (Kamerstukken II 2010/11, 32549, nr. 2) of the Dutch Telecommunication Act. It includes the implementation of Article 5(3) and Article 6(3) of the ePrivacy Directive 2002/58/EC into national law. In a nutshell, informed consent has to be achieved before data that is stored in the terminal equipment of an individual, may be collected and processed. The *Besluit universele dienstverlening* *en eindgebruikerbelangen (BUDE)* [vEZ04] deals also with informed consent in Article 4(1). According to Article 4(1), clear and comprehensive information has to be given and individuals must have the right to refuse (i.e. opt-out). This ”Besluit” will thus have to be amended in accordance with the new law, probably that amendment is already in the making.

An ongoing point of discussion was whether to add the word *unambiguous* before the word *consent* (Kamerstukken II 2010/11, 32549, nr. 14) in the amended act. Finally, the word was not added and the Act was amended accordingly. The amended Act is now in the First Chamber. In the amendment, Article 11.7a is changed that deals with tracking and behavioral advertising.

> *”2. Aan het eerste lid wordt een volzin toegevoegd, luidende: Een handeling als bedoeld in de aanhef, die tot doel heeft gegevens over het gebruik van verschillende diensten van de informatiemaatschappij door de gebruiker of de abonnee te verzamelen, combineren of analyseren voor commerciï¿½le, charitatieve of ideï¿½le doeleinden, wordt vermoed een verwerking van persoonsgegevens te zijn, als bedoeld in artikel 1, onderdeel b, van de Wet bescherming persoonsgegevens.” (Kamerstukken II 2010/11, 32549, nr. 39)*

The Dutch Data Protection Act (Wet Bescherming Persoonsgegevens) is thus always applicable when users are getting tracked for commercial purposes. Cookies may only be placed on the terminal equipment of a user after clear and comprehensive information is provided and consent is given. Consent cannot be given by using the appropriate browser settings because, according to the secretary of state (Kamerstukken II 2010/11, 32549, nr. 2), most users are not aware of the fact that most browsers accept all cookies by default. This could change in the future when browsers offer the ability to specifically indicate which party or website may place cookies. With regard to behavioral advertising, cookies may only be placed after consent is given and the Dutch Data Protection Act is always in effect. Therefore, consent is not only necessary for placing or reading of cookies but also for collecting and processing of personal data. Thus, if a user has given his consent for placing a cookie does not necessarily mean that he has also given his consent for the collection of personal data. In the amendment is stated that the Dutch Data Protection Act is always applicable when users are getting tracked. If the amendment will be accepted, then advertising networks have to comply with what is stated in the Dutch Data Protection Act. From that moment on, cookies for behavioral advertising may only be placed after consent is given regardless of what kind of data is collected. Furthermore, the whole behavioral advertising technique has to be redesigned because the new law asks for an opt-in policy. Therefore, consent by browser settings is not sufficient. The Dutch Tweede Kamer[2] also recognized during the vote on the amendment of the Dutch Telecommunication Law that with today's browsers it is not possible to achieve informed consent. They suggest that browser vendors start working on new versions that comply with what is stated in the new law.

Mark Jansen, a Dutch advocate, says that it makes no sense to turn all attention on cookies[3]. According to him, it would make more sense to change existing laws in such a way that all tracking techniques are covered. Regulating tracking technologies and, thus, behavioral advertising, would solve many privacy related concerns.

3.7 Recent Achievements and Unsolved Problems

In this section, we will deal with recent achievements and unsolved problems regarding data protection and privacy in the EU. Some examples are given in both sections. We will also discuss how countries that are not member of the EU deal with data protection and privacy.

3.8 Recent Achievements and Unsolved Problems

In this section, we will deal with recent achievements and unsolved problems regarding data protection and privacy in the EU. Some examples are given in both sections.

3.8.1 Achievements

The European Union Agency for Fundamental Rights published in June 2011 a rapport about challenges and achievements in 2010[4]. There is a need to review and modernize the current EU data protection framework because new technologies raise fundamental rights concerns and the current EU data protection framework is already outdated. Technology neutrality seems to be an important issue.

The Dutch Telecommunication Act is recently revised by the Dutch Tweede Kamer. A sentence was added to the first paragraph:

> *Een handeling als bedoeld in de aanhef, die tot doel heeft gegevens over het gebruik van verschillende diensten van de informatiemaatschap-*

[2] http://www.eerstekamer.nl/9370000/1/j9vvievljta5h6f/viqfiqqnfc3t/f=y.pdf (accessed 24-08-2011)

[3] http://dirkzwagerieit.nl/2011/02/10/liggen-cookies-de-wetgever-terecht-zo-zwaar-op-de-maag/ (accessed 24-06-2011)

[4] http://fra.europa.eu/fraWebsite/attachments/annual-report-2011_EN.pdf (accessed 23-08-2011)

> *pij door de gebruiker of de abonnee te verzamelen, combineren of analyseren voor commerciële, charitatieve of ideële doeleinden, wordt vermoed een verwerking van persoonsgegevens te zijn, als bedoeld in artikel 1, onderdeel b, van de Wet bescherming persoonsgegevens. (Kamerstukken II 2010/11, 32549, nr. 39)*

In the revised Telecommunication Act is clearly stated that data that is collected and processed for commercial purposes is protected by the Dutch Data Protection Act (Wet Bescherming Persoonsgegevens). By adding the sentence it is now unambiguously clear that users must have the right to access their data.

The Unabhängige Landeszentrum für Datenschutz (ULD), the Data Protection Authority (DPA) from the state of Schleswig-Holstein (Germany) published in a press release[5] in September 2011 that they want to remove all Facebook Like Buttons (see section 2.4.3 from their pages.

> *"As long as Facebook cannot provide information about the purposes for which and what kind of traffic data are processed in the USA, website owners in Germany cannot answer for the passing on of these data."*

The ULD wants to achieve that Facebook complies with existing data protection law.

Viviane Reding, European Commission Vice-President said on 11 January 2011:

> *"I would like to single out (. . .) priority areas where I believe we need to show strongly that Europe's policy is changing with the Lisbon Treaty. First of all, we need to strengthen substantially the EU's stance in protecting the privacy of our citizens in the context of all EU policies."*

Reding recognizes that EU data protection legislation needs to be modernized.

Goldfarb and Tucker [GT11] did research on the effectiveness of advertisements in and outside the EU. One of their findings is that *"in Europe, where privacy laws have been implemented, banner ads have experienced a reduction in effectiveness of 65% on average in terms of changing stated purchase intent" [GT11].* Non-EU countries did not show the same changes in the same time frame. Thus, privacy regulations can have an impact on the effectiveness of behavioral advertising but we have to remark that the results

[5] `https://www.datenschutzzentrum.de/presse/20110930-facebook-enforce-privacy.html` (accessed 04-10-2011)

are based on the outcomes of one experiment. The experiment consisted of a survey asking people whether they changed their intent, this may not be very accurate. Furthermore, websites, such as news sites, have a lot of different content. This makes it more difficult to link appropriate advertisements to those sites and, thus, without behavioral advertising, individuals will more frequently see advertisements that do not match with their interests. Hence, they conclude that only websites with a specific content that do not use behavioral advertising can still be easily targeted with corresponding advertisements. Finally, Goldfarb and Tucker [GT11] state that the regulations could lead to more obtrusive advertisements to get the attention of the individual.

The Data Protection Directive 95/46/EC was enacted in 1995. In the last 16 years, a rapid evolution of the information society took place and exchange of data is, now, a common practice. The data protection framework must be capable of dealing with technologies that gather huge amounts of data every day. Also the Data Retention Directive 2006/24/EC is under review because there are concerns that the directive does not comply with fundamental rights of the European Union. Data retention forms a risk when data is being leaked or abused. Finally, technological challenges such as new location-based services, Facebook and Google Street View are discussed in terms of privacy concerns. All three can also be used for commercial purposes.

3.8.2 Unsolved Problems

According to Kuner [Kun08], the laws have to be more transparent for both data controllers and citizens so that everyone knows their rights and obligations. Kuner also mentions in his book *European Data Protection Law, Corporate Compliance and Regulation* [Kun07] that it is difficult to determine whether personal data is really anonymous. Data protection authorities have to examine each case in detail before it can be stated that the data cannot be traced back to a single person. ISPs, for example, can determine the identity of each customer when only the IP address is known. A question that arises is whether the IP address can be seen as a pseudonym? Advertising networks that use the IP address to track users, collect data and generate profiles, must give the user the right to access in case that the IP address is considered to be a pseudonym. The legal consequence of being a pseudonym is that it must be treated as personal data. The same holds for browser fingerprints because they are also unique and can be considered as a pseudonym.

Kuner [Kun07] further mentions some problems with achieving informed consent. According to him, standard terms and conditions are often very long and not transparent. Thus, data subjects often accept them without fully understanding the content. The interdependence of the rules and obligations for data controllers and data processors are sometimes ambiguous. Kuner [Kun07] says that there is a lack of clarity between the roles. Data controllers must take most of the necessary measurements whereas data processors have to carry out the instructions of the data controller. Sometimes both roles overlap and it becomes difficult to determine whether a party is a controller or processor. Data controllers can also process data on behalf of another data controller. If this is the case, then they are also data processors. It is also possible that two organizations have the same data source and organization A processes also data for organization B and the other way around. In case of behavioral advertising, it would be difficult to assign responsibilities and obligations.

Curren and Kaye [CK10] say that individuals must be more in control of their own data which would have a positive effect on the awareness of an individual. Therefore, individuals must be provided with methods that put them into control. In case of behavioral advertising, users should have access to the profiles they match. These methods should also support an effective right of withdrawal of consent and clear guidance should be given to those who have to implement them.

TNO, a Dutch research institute, published in February 2011 a rapport [TI11] about problems with regard to the application of the Data Protection and ePrivacy Directive in case of behavioral advertising. The central problem according to [TI11] is that most parties that have to comply with what is stated in the directives are not well informed. Furthermore, TNO argues that there is often a lack of knowledge on both sides, data controllers and data subjects. This is also noted by Kuner [Kun07] about the data subject. They also question if the existing law gives clear guidelines. One of the findings of the TNO survey is that responsibilities are hardly divided between the different parties. Agreements about who is responsible for the provision of information simply do not exist. It is thus important to clearly define roles and responsibilities. Another point of discussion is the terminal equipment that is used by more than one data subject. Therefore, consent has to be given by different data subjects which makes it even more difficult to find an appropriate solution. TNO also mentions that placing information in privacy policies and general terms of agreement is not sufficient but most advertising networks have the opinion that this is consistent with the directives.

Kuner [Kun08] proposes to combine both directives *into a single instrument.* This would make it more understandable and the Working Party and European Commission could publish guidelines and case studies. Furthermore, he says that all parties (citizens, data protection authorities and data controllers) should be involved.

Justice Commissioner Viviane Reding promoted in a speech[6] in March 2011 four pillars:

1. the right to be forgotten
2. transparency
3. privacy by default
4. data protection regardless of data location

According to Reding, the right to privacy should be built on these pillars.

The European Commission adopted in November 2010 a strategic communication [Com10] that deals with ideas on how to modernize the current data protection framework. It could also be possible to integrate both directives into a single instrument. In paragraph 2.2.4 of the rapport the responsibilities of data controllers is discussed. According to the Commission, the obligations of data controllers should be more clearly in the framework and data protection supervisory authorities should be involved. Furthermore, the same responsibilities should also be applied to data processors. With regard to behavioral advertising, privacy enhancing technologies and privacy by design should be promoted. Finally, a data protection impact assessment should be carried out in cases when specific technologies such as profiling that involves specific risks are used.

[6] http://europa.eu/rapid/pressReleasesAction.do?reference=SPEECH/11/183&format=HTML&aged=0&language=EN&guiLanguage=en (accessed 13-09-2011)

Chapter 4

Privacy-enhancing Technologies for Behavioral Advertising

This chapter deals with privacy-enhancing technologies (PETs). We first give a general introduction and make a clear differentiation between privacy-enhancing technologies and transparency-enhancing technologies (TETs). Some of the PETs discussed in this chapter are also dealing with transparency. We will not discuss different TETs in detail because they are mostly integrated in PETs. Finally, we analyze five existing PETs and discuss shortcomings. We will introduce our own approach in chapter 5.

4.1 Introduction

Privacy-enhancing technologies (PETs) are getting more attention because companies, such as advertising networks, collect (personal) data every day and consumers are often not aware of what is happening in the background when they connect to the internet. We will use the following definition by van Blarkom et al.:

> *Privacy-Enhancing Technologies is a system of ICT measures protecting informational privacy by eliminating or minimising personal data thereby preventing unnecessary or unwanted processing of personal data, without the loss of the functionality of the information system. [vBBO03]*

Thus, PETs focus rather on protecting consumer privacy than making it more transparent for consumers. A typical PET can be described as a firewall that protects the (personal) data of a consumer.

Transparency-enhancing technologies (TETs) differ from PETs in way that they focus on making data collection and processing more transparent for the consumer. In FIDIS Deliverable 7.7 TETs have been defined as:

> *Transparency enhancing technologies, which have not been developed yet. Their function is not the history management of the data of a data subject, but the anticipation of profiles that may be applied to this particular subject. This concerns personalised profiles as well as distributive or non-distributive group profiles, possibly constructed out of anonymous data. The point would be to have some idea of the selection mechanisms (application of profiles) that may be applied, allowing a person adequate anticipation. To be able to achieve this the data subject needs access - in addition to his own personal data and a profiling / reporting tool - to additional external data sources, allowing some insight in the activities of the data controller. Based on this additional information the data subject could perform a kind of counter profiling. [WHM06]*

In FIDIS Deliverable 7.12 [WH09] Hildebrandt et al. argue for TETs that provide transparency about the profiles a user matches. The focus is not personal data, but any relevant profile. There are PETs, focused on data minimization, informed consent etc. and there are TETs that provide transparency about the purpose of processing and the processing of personal data. These are often seen as a subcategory of PETs because they focus on personal data. We can state that TETs are needed rather than PETs because to make an informed choice about the use of PETs users require TETs that show what profiles they match.

The Federal Trade Commission (FTC), an agency of the US government that deals with consumer protection, published in 2010 the report *Protection Consumer Privacy in an Era of Rapid Change* [FTC10]. FTC recognizes the privacy concerns of consumers and, therefore, proposes a framework for business and policy makers. The framework consists of three principles:

1. Privacy by design
2. Simplified choice
3. Greater transparency

Privacy by design means that privacy has to be part of every development stage of a new product or service. Companies must implement privacy protecting measures. Consumer choice should be simplified in a way that the consumer decides if he or she wants to share (personal) data. The last principle, greater transparency, supports TETs because it asks for more transparent data practices. Standards for privacy terms should be developed. It must be possible to compare different privacy practices in an easy way. These goals can only be achieved when all stakeholders work together.

Also Koorn and ter Hart mention in their paper [KtH11] that privacy by design should be a leading principle. Organizations often collect more data about their customers than that is actually needed for the performance of a contract. Limiting data collection and processing by privacy by design lowers the risks. Therefore, Koorn and ter Hart say that one important issue is to determine if the processing of personal data is necessary, necessary to a limited extent or avoidable. This is already the case when the Data Protection Directive 95/46/EC is applicable. Article 6 and 7 of the directive deal with fair processing, lawful grounds and the data minimization principle.

Goldberg [Gol] examined the development of different PETs in the last years. He experienced that numerous PETs do not make it out of the lab. They do not improve the security and privacy of users at all. Therefore, Goldberg has identified four properties that useful security and privacy enhancing technologies must have:

- Usability
- Deployability
- Effectiveness
- Robustness

PETs must be easy to use. Users that get frustrated from a PET, simply turn it off. Furthermore, PETs must be compatible with the user's software. The ideal idea is to implement PETs in already existing software. PETs that have to be installed separately are less preferred because users have to install them manually. PETs must be effective and protect the privacy of the users. Open design and implementation can help because everyone can review it and security flaws can be detected much easier. Furthermore, robustness means that a PET must be able to operate even during some kind of disturbances, for example, when the computer of a user contains worms or viruses. According to Goldberg, developers should always keep these principles in mind.

Camp and Osorio [CO03] examined different PETs in 2002. They use three concepts of privacy: autonomy (watched people are not free), seclusion (right to be left alone) and property (data are valuable). Their findings are that most privacy enhancing providers do not well understand the privacy market. Introducing new PETs seems to be difficult because, even when the software has a reduced price, most users are not willing to install it. Users are cautious when, for example an ISP, tells them to install software that protects personal data because they often do not know what really happens with their data.

4.2 Approaches

This section deals with five different PETs. The Stanford Center for Internet and Society has a list[1] with more than 50 PETs. Some of them, such as AdBlock, got popular in the last years but the majority remains scientific approaches. There are also numerous EU funded projects such as Prime[2] (Privacy and Identity Management in Europe), FIDIS[3] (Future of Identity in the Information Society) or PrimeLife[4] (Privacy and Identity Management in Europe for Life). These projects focus mainly on identity management for social network sites and sale of goods and services whereas our research domain is about tracking and profiling with regard to behavioral advertising. In the following sections, we will introduce third-party cookie blocking, Do Not Track, Track Me Not, P3P and Tracking Protection Lists. We choose these five PETs because they got popular in the last years and are supported by most browser vendors.

4.2.1 Third-party Cookie Blocking

How It Works

Most browsers give users the choice to block third party cookies by default. When a user decides to turn third party cookie blocking on, cookies from websites that are not directly visited are blocked. Many websites contain third party content from advertising networks but we have to remark that not all types of tracking technologies are blocked when third party cookies are disabled. For example, content (e.g. flash cookies) can still be stored outside the domain of the browser without the knowledge of the browser.

Evaluation

Third-party cookie blocking has the following characteristics:

- Opt-out solution
- Compliance checking by examining browser cookies
- Transparency is not given

[1] http://cyberlaw.stanford.edu/wiki/index.php/PET (accessed 26-07-2011)
[2] https://www.prime-project.eu/ (accessed 23-08-2011)
[3] http://www.fidis.net/ (accessed 23-08-2011)
[4] http://www.primelife.eu/ (accessed 23-08-2011)

- User is in control

It is a typical opt-out solution because it is not turned on by default in most browsers. Compliance can only be checked by examining if cookies from advertising networks are installed. Browser vendors do not guarantee that all third-party cookies are blocked. It is, thus, still possible that user are getting tracked even when third-party cookie blocking is enabled. Browser vendors do not make third-party cookie blocking for users transparent. The check box for turning on/off third-party cookie blocking is buried in the privacy settings of a browser. In case that third-party cookie blocking is disabled, it is not transparent for the user. It is unknown to the user which advertising networks are tracking him/her and what kind of information the profile contains. Finally, we have to remark that the user is in control because the browser blocks unwanted, third-party cookies.

4.2.2 Do Not Track (DNT)

How It Works

Do Not Track[5] is a technology that offers to user the possibility to opt out of tracking by websites, such as advertising networks, they do not visit. Researchers from Stanford University created the project. There is an ongoing discussion about making Do Not Track a standard [FTC10]. The idea behind Do Not Track is that it adds a message to the HTTP header when a HTTP request is made [MNS11]. The header already exists and an extra field, DNT, is added that indicates whether someone wants to be tracked or not. Thus, the technology is an opt-out solution. In figure 4.1 is illustrated what the three parties, user, website and advertising network, must do or comply with.

First, the user must decide whether or not he/she wants to be tracked by advertising networks in the privacy settings of the browser. The browser will add the DNT preference to each HTTP request when a website is visited or third party content is requested. Thus, the DNT preference is also transmitted to third parties, such as advertising networks, when advertisements are being loaded. We have to remark that for any content that has to be delivered to the user a HTTP request must be made first. Thus, the DNT preference will always be transmitted to the server of the website and to servers of third parties when a request for third party content is made. The advertising network receives the preference of the user and must comply with it, otherwise Do Not Track is not effective.

[5] `http://donottrack.us/` (accessed 07-07-2011)

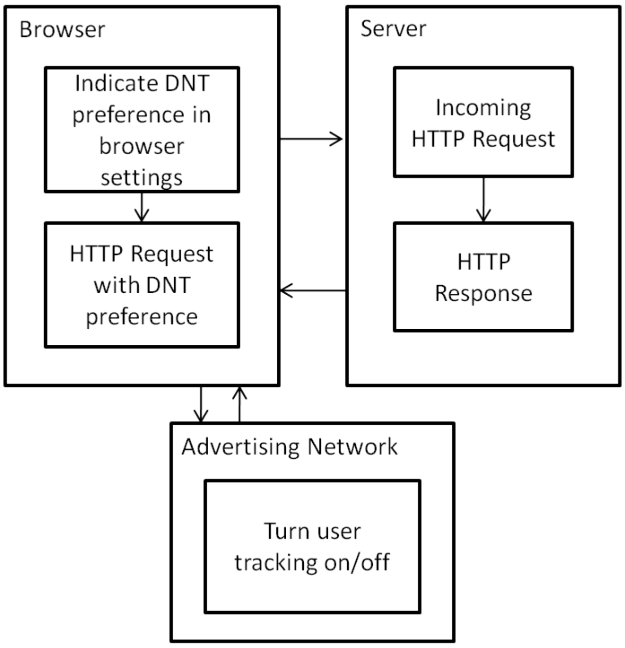

Browser
Indicate DNT preference in browser settings
HTTP Request with DNT preference
Server
Incoming HTTP Request
HTTP Response
Advertising Network
Turn user tracking on/off

Figure 4.1: Do Not Track

Evaluation

The developers of Do Not Track want to achieve that users can easily opt-out independent from the tracking method that is used[6]. DNT has the following characteristics:

- Opt-out solution
- Compliance checking is difficult
- Transparency is not given
- Advertising network is in control

The success of this technology depends on the willingness of browser vendors to support Do Not Track and the willingness of advertising networks to comply with the users wishes. An existing problem is how to deal with tracking technologies, such as flash cookies, that are locally stored because the browser cannot access them. The latest versions of Firefox[7], Internet

[6] http://www.pcworld.com/article/223633/the_state_of_do_not_track_on_the_internet.html (accessed 07-07-2011)

[7] http://dnt.mozilla.org/ (accessed 07-07-2011)

Explorer[8] and Safari[9] support Do Not Track. Google comes up with their own implementation of Do Not Track for Google Chrome. They call it ChromeBlock and, according to Google, users can choose which advertising network they want to shut down[10]. DNT can also be easily changed to an opt-in solution when tracking by default is prohibited and the DNT header indicates that someone wants to be tracked. Compliance can be checked by examining the installed cookies. The advertising network is in control because it decides if a cookie will be sent to the user. Even when the user wants to be tracked it is still not comprehensible what kind of information is collected and for what purpose.

Tschofenig and van Eijk [TvE] argue that DNT can only be successful when advertising networks are forced by law to comply with the wishes of the user. The Federal Trade Commission proposed a framework for business and policy makers [FTC10] but it is still unclear what the lawmaker is going to do with that. The difference between standard third party cookie blocking and Do Not Track is that in case of standard third party cookie blocking most third party cookies will be blocked. There is no differentiation between cookies that are tracking cookies and cookies that are just third party cookies. In case of Do Not Track it is possible to differentiate between normal and tracking cookies but, as already stated, this depends of the willingness of advertising networks to comply.

4.2.3 Track Me Not (TMN)

How It Works

Howe and Nissenbaum [HN09] developed Track Me Not. TMN is PET that obfuscates users internet behavior by adding random queries. It is difficult for advertising networks that track users to differentiate between queries that are made by the user and those that are made by TMN. TMN has the following technical mechanisms [HN09]:

- dynamic query lists
- real-time search awareness
- live header maps
- burst-mode queries
- selective click-through

[8] `http://www.msnbc.msn.com/id/40554324/ns/technology_and_science-security/t/microsoft-unveils-do-not-track-ie-feature/` (accessed 07-07-2011)

[9] `http://online.wsj.com/article/SB10001424052748703551304576261272308358858.html` (accessed 07-07-2011)

[10] `http://google-chrome-browser.com/tags/do-not-track` (accessed 07-07-2011)

Dynamic query lists contain popular websites and popular search terms that are updated by RSS. With real-time search awareness, TMN gets to know when a user starts a search query and can use the information for further queries. TMN uses live header maps to reproduce the exact headers that were used the last time when a user made a request. Differentiations between user and TMN requests cannot be made. Burst-mode queries are several context dependent requests that are made in a short time period. The behavior of TMN is more cloth to actual user behavior. Finally, selective click-through opens additional links on a results page.

Evaluation

TMN has the following characteristics:

- Opt-out solution
- Compliance is not necessary
- Transparency is not given
- Users is in control

TMN is not a typical opt-out technology because it allows advertising networks to track the user but the user is still in control. By obfuscating the online behavior makes profiling almost impossible because the profiles are getting too general and not specific enough. It is thus hard to examine which search terms or visited websites really reflect the interests of a user. Advertising networks do not have to comply but we have to be critical about the success rate of TMN. Is it really impossible to extract the user profile out of the TMN profile or do repeating patterns exist?

4.2.4 P3P

How It Works

The World Wide Web Consortium (W3C) recommended in 2002 the Platform for Privacy Preferences Project (P3P). We will examine the specifications in [CLM$^+$02]. P3P is a standard format for privacy practices. It can be implemented in web sites and user agents can automatically interpret it. According the W3C, one of the benefits of P3P is that users do not have to read the privacy policies of every web site they visit. Furthermore, P3P offers the possibility to inform users before personal data is being transferred to the web site. A P3P policy includes the following:

- P3P base data schema

- Standard set of uses
- XML format
- A means of associating privacy policies
- A mechanism that enables the transport of P3P over HTTP

Web sites define in the P3P base data schema the data they want to collect. In the standard set of uses, data categories and privacy disclosures are defined. All information is expressed in XML and is transported over HTTP. P3P has two goals: first, web sites can present their data collection practices in a standardized manner. Second, users will be informed about what data will be collected, how the data will be used and the possibility to opt-out or opt-in. Before a HTTP transaction with P3P can be made, both web site and user have to define their own P3P policy. The website expresses in their P3P policy which kind of data will be collected and for what purposes. Furthermore, information about the person or organization that collects the data is given and with whom the data is being shared. The user defines his/her own preferences and expresses which data may be shared for what purposes. A simple HTTP transaction including P3P is illustrated in figure 4.2.

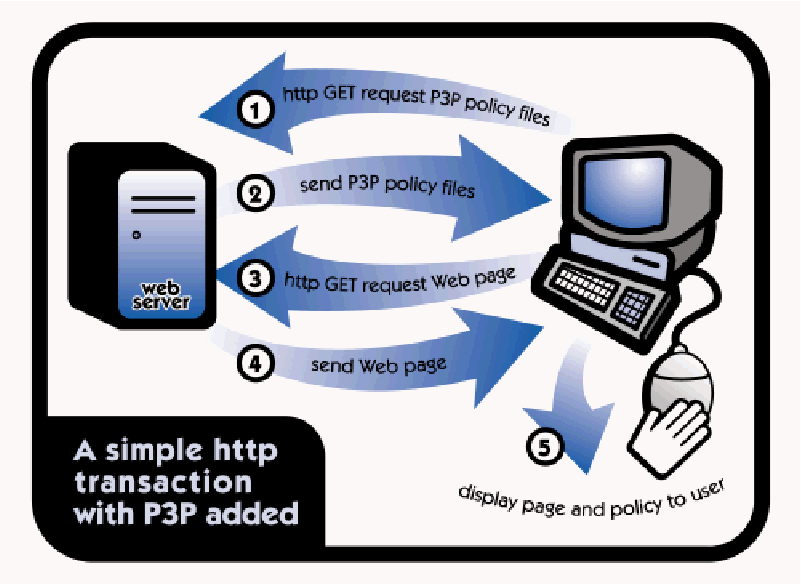

1 http GET request P3P policy files
2 send P3P policy files
3 http GET request Web page
4 send Web page
5 display page and policy to user
web server
A simple http transaction with P3P added

Figure 4.2: P3P transaction[11]

When a user wants to visit a certain website that supports P3P, first the P3P policy of the web server is transmitted to the user. The P3P user client

[11] http://www.w3.org/2000/07/transaction.png (accessed 05-05-11)

then compares the web server policy with the user policy and, if the policy is accepted by the user client, the content of the website will be loaded.

Evaluation

P3P has the following characteristics:

- Opt-out/Opt-in solution
- Compliance checking is difficult
- Transparency by policies
- Advertising network is in control

P3P is not a typical opt-in or opt-out solution. It depends on what is stated in the policy of the user and of the website. Compliance checking seems to be difficult. The W3C mentions in the specifications [CLM+02] of P3P that it is not possible to control whether web sites act according to what they state in their P3P policies. The user cannot verify if what is stated in the policy is really true. P3P offers far more transparency to its users than third-party cookie blocking or Do Not Track. Users can define what kind of data they want to share and for which purpose but we have to mention that the advertising network is still in control because tracking cookies can still be sent.

The policy exchange is mainly taking place between websites and users. Third parties, such as advertising networks, will only be involved when they also comply with P3P. Thus, websites that contain advertisements could state that they do not collect data themselves but, in fact, they support data collection by third parties. Users get only a notification about whether a web site uses P3P or not and, in case the site does, information about what is stated in the policy.

We have to mention that P3P will only be successful when it is supported by web site owners, browser vendors and networks that deliver third party content. There are also criticisms who say that P3P is a *"complex and confusing protocol"*[12]. It is not user-friendly and even misleading because it facilitates data collection. Web sites that are mainly interested in collecting data of their visitors could state that they only collect some anonymous data that is stored in web logs but, in fact, collect far more data. Users who have given consent to the collection of anonymous data in their P3P preferences visit the site without knowing that additional information is collected as well. Furthermore, it cannot be guaranteed that the processing of data complies with Article 6 Data Protection Directive 95/46/EC. P3P could be

[12] `http://epic.org/reports/prettypoorprivacy.html` (accessed 12-07-2011)

a tracking technology independent solution when all parties comply with it but, as already stated, it can also easily be misused.

4.2.5 Tracking Protection List (TPL)

How It Works

Tracking protection lists contain URLs of websites that often deliver third party content. A common definition of a tracking protection list is the following:

> *A Tracking Protection List (TPL) contains web addresses (like msdn.com) that the browser will visit (or "call") only if the consumer visits them directly by clicking on a link or typing their address. By limiting the calls to these websites and resources from other web pages, the TPL limits the information these other sites can collect.*[13]

Web addresses of advertising networks that are listed will not be requested when someone visits a web site that contains third party content. In figure 4.3 is illustrated what a tracking protection list does.

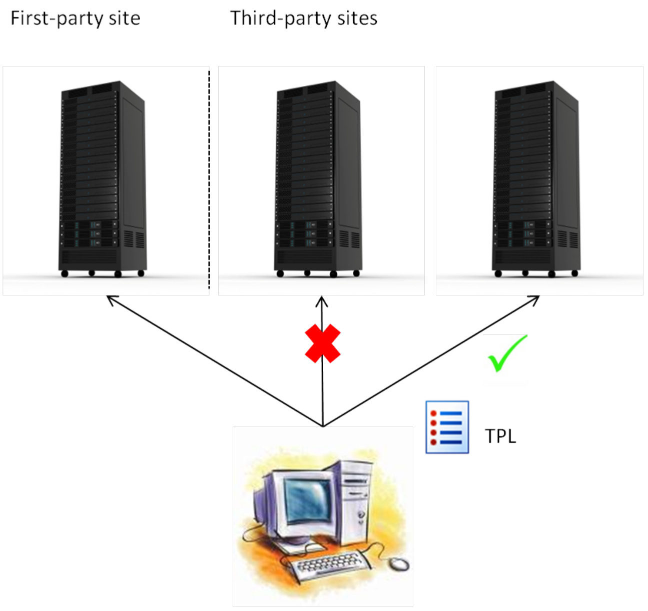


Figure 4.3: Tracking Protection List Schema

[13] http://blogs.msdn.com/b/ie/archive/2010/12/07/ie9-and-privacy-introducing-tracking-protection-v8.aspx (accessed 12-07-2011)

Evaluation

A TPL has the following characteristics:

- Opt-out solution
- Compliance checking is not necessary
- Transparency is not given
- User is in control

TPL is a typical opt-out solution. Users can control themselves which third party sites are blocked. When we assume that a tracking protection list exists which covers all third party domains of advertising networks, then user tracking with cookies or web beacons would not be possible because the browser checks if the HTTP address is listed before the HTTP request is sent. Only the first party site is loaded by the browser without third party content. Thus, no communication between the browser of the user and third parties takes place. Thus, checking whether an advertising network complies is not necessary. Third-party sites that are not listed in the TPL can still send cookies. Transparency for those sites is not given because the user does not know what is going to happen with the data that is collected about him/her. Thus, the user in control about known advertising networks in the sense that he/she can decide which advertisements, and thus cookies, are accepted or rejected. Tracking and profiling by advertising networks that are listed will not be possible with the common technologies.

Microsoft implemented tracking protection lists in Internet Explorer 9[14]. We have to remark that tracking protection lists must be updated regularly and be published by trusted authorities. Finally, there is no guarantee that all third party domains will be listed because web addresses can change very often and new web sites come up every day.

4.3 Discussion

A *Study on the economic benefits of privacy-enhancing technologies (PETs)*[15] by London Economics shows that benefits of PETs are technology-specific and dependent on the applications in which they are used. The conducted survey shows that individuals are concerned about privacy. Individuals are often uncertain about the risks and do not have sufficient knowledge about

[14] http://www.microsoft.com/presspass/features/2010/dec10/12-07ie9privacyqa.mspx (accessed 12-07-2011)

[15] http://ec.europa.eu/justice/policies/privacy/docs/studies/final_report_pets_16_07_10_en.pdf (accessed 01-09-2011)

PETs. Furthermore, without legislation that creates a level playing field there is no incentive for the industry to invest in PETs. We will now discuss shortcomings of the five earlier discussed PETs with regard to our research question.

4.3.1 Opt-In/Opt-Out

In general, we can conclude that all PETs do not satisfy the requirements of the European Union because all are opt-out solutions. Only P3P can also function under certain circumstances as an opt-in PET but this is only possible when advertising networks do not track users by default and comply with what they state in their policy. Prior informed consent can only be achieved with opt-in technologies. Do Not Track and P3P are highly dependent on the willingness of advertising networks to satisfy the users wishes by implementing the technology in their databases and tracking methods. Tracking protection lists only opt-out from already known advertising networks. Unknown advertising networks can still track users even when a Tracking protection list is used. Third-party cookie blocking does not block all kinds of third-party cookies and, thus, users are still getting tracked. Track Me Not focuses not on blocking tracking mechanisms but tries to opt-out by obfuscation. It seems difficult to examine whether a user really opts-out when TMN is used.

4.3.2 Compliance

In case of third-party cookie blocking, Tracking Protection Lists and Track Me Not compliance by advertising networks is not necessary because the user controls whether or not he or she wants to share his or her data. Tracking mechanisms, such as cookies and web beacons, are blocked or advertisements are not being loaded (TPL). Furthermore, users can control if tracking cookies are installed by examining the list of cookies in their browsers. Only in case of Track Me Not data is still be shared together with random data. However, compliance is still not necessary because it is hard to examine for advertising networks which part of the data belongs to the user and which part is random data.

In case of Do Not Track and P3P compliance is uncertain because the advertising network is in control and decides whether or not to track a user by transmitting cookies or web beacons. Thus, it depends on the willingness of advertising networks to stop data collection when a user indicates that he or she does not want to be tracked. The list of cookies can be examined as well. P3P can also easily be abused to collect large amounts of data when web sites claim in their policies that they only collect, for example traffic data. If we assume that many users agree with the collection of traffic data, then additional data could be collected without knowledge of the users.

4.3.3 Transparency

In case of opt-out, all five PETs do not have a sufficient level of transparency. Most users are not familiar with cookies and other tracking methods and, thus, cannot control if advertising networks collect data of them even when they opt-out. Also the list of cookies is not easily accessible because it is hidden in the privacy settings of a browser. Furthermore, it is difficult for unexperienced users to differentiate between normal and tracking cookies.

In case of opt-in, P3P is most sophisticated. User can define their own privacy policy. For all five PETs except P3P holds that the processing of data and, thus, the generation of profiles is not transparent. Also the purpose of data processing remains unclear. According to W3C, users who use P3P are getting informed about what data is collected and how the data is being used. It seems unclear how this is taking place in practice. For users, it is not only important to know what data and how it is collected but also how the resulting profiles look like. Thus, informed consent cannot be achieved when a user does not know which profiles he/she matches.

Summarized, we can say that all five PETs do not meet the requirements of EU legislation, notably Article 5(3) and 6(3). A point of discussion is whether a PET can meet the requirements of the EU without the collaboration with advertising networks and browser vendors. It seems not possible to protect the privacy of individuals by just using a PET. Transparency with regard to personal data and the profiles they match must be guaranteed.

Chapter 5

Achieving Informed Consent

In the first part of this chapter we give recommendations with regard to the modernization of the data protection framework. In the second part we describe a privacy-enhancing technology for behavioral advertising that fulfills the requirements of EU legislation. Finally, we will give an answer to our research question.

5.1 The Legal Aspect

5.1.1 Beyond Personal Data

First, we have to define which data will be covered by our approach. The scope of the Data Protection Directive 95/46/EC is limited to *personal data*. The scope of the ePrivacy Directive 2002/58/EC is broader because the term *any data* is used. We have to remark that this leads often to ambiguous situations because advertising networks and national law makers do not fully understand the meaning and relation between both terms. Therefore, it is necessary that the Data Protection Directive 95/46/EC is always applicable when data is collected and processed for behavioral advertising purposes. Grey areas in law will remain when the definition leaves space for various and ambiguous interpretations. In our approach, we will not differentiate between personal or non-personal data with regard to behavioral advertising. In case of behavioral advertising, it makes no sense to differentiate between certain types of data because the consumer's privacy has to be protected under all circumstances and profiling can have substantial consequences. Therefore, we recommend that 5(3) and 6(3) concern all types of tracking and tracing. In the amendment (Kamerstukken II 2010/11, 32549, nr. 39) of the Dutch Telecommunication Law all types of data with regard to behavioral advertising are protected by the Dutch Data Protection Act. When advertising networks are not longer successful in arguing that their kind of data collection and processing forms an exception then we are a step further in protecting the privacy of consumers. We have to mention that the Dutch law already declares that all data involved in tracing and tracking for the purpose of profiling in a commercial or charity context will be considered personal data in the sense of the Dutch Data Protection Act

5.1.2 Informed Prior Consent

Our approach is based on an opt-in policy. It is not allowed to collect any data from any person for advertising purposes before consent is given. We recommend that no exceptions may exist because that would lead to the old situation where it is a common practice to achieve consent indirectly. We suggest that consent has to be given directly by clicking on a button. We will go into detail in section 5.2. Without explicit consent, only contextual advertising should be allowed because only information about the content of the website is used to deliver appropriate advertisements. The ePrivacy Directive 2002/58/EC supports the idea of informed consent before someone is allowed to collect information but they also say that this can be achieved, under certain circumstances, by using the appropriate browser settings (see section 3.4.2). We do not support this idea because, according to our opinion, the whole idea of informed consent depends on whether a user is aware of what is happening with his/her (personal) data and, in case of behavioral advertising, what profiles he/she matches. Awareness cannot be guaranteed by configuring the browsers privacy settings once. Thus, no exceptions for achieving informed consent should be allowed.

5.1.3 Two Directives: One Goal?

In section 3.8.2, we mentioned that Kuner [Kun08] proposes to combine the Data Protection Directive 95/46/EC and the ePrivacy Directive 2002/58/EC into a single instrument. Does this really solve the problem? In our opinion, just combining both directives will not solve the entire behavioral advertising problem. Most of the privacy concerns will still remain. Therefore, it seems necessary to work on a new directive that complements or amends the Data Protection Directive 95/46/EC and the ePrivacy Directive 2002/58/EC. In the following, we will discuss problems that have to be solved by the new directive.

We already discussed the problem of defining the term data in section 5.1.1. The scope of the new directive should be any kind of data collection or processing for commercial purposes. The Data Protection Directive 95/46/EC should always be applicable when tracking technologies are used for the collection of data and group profiling. Furthermore, the new directive should recommend clear and comprehensive opt-in technologies. The different opt-in technologies should not support automated decision making because this can be compared with letting the browser make the decision. Therefore, all tracking technologies, and thus behavioral advertising, should be turned off by default. This will also solve problems with regard to achieving informed consent in section 5.1.2 because exceptions that can be used by advertising networks for collecting data will no longer exist. Also the identity of the controller and, if so, of processors and third parties must be provided to the user. In our opinion this can be achieved by introducing a new standard for this kind of information. It can be, for example, a simple XML schema that covers all the necessary information. Introducing a standard has the advantage that browser vendors can always interpret the received information of advertising networks and then show it to the user. The same holds for the problem concerning the knowledge of processing.

We already mentioned that users must be aware of what kind of information will be processed. We suggest that advertising networks have to provide access to every profile that a user matches after a user has already given his consent. A point of discussion is how detailed a profile must be described. We suggest that at least the categories that someone matches and the possible categories of interest should be shown to the users. The German Federal Data Protection Act (BDSG)[1] explicitly deals with profiling in Article 4(e) and Article 34(2):

> **Section 4e Contents of notification**
> Where automated processing operations are subject to the obligation to notify, they shall include the following information:
> [. . .]
> 5. a description of the category or categories of data subject and of the data or categories of data relating to them
> [. . .]
>
> **Section 34 Access to data**
> (2) In the case of Section 28b, the body responsible for the decision shall provide information to data subjects at their request concerning
> 1. probability values calculated or recorded for the first time within the six months preceding the receipt of the information request,
> 2. the types of data used to calculate the probability values, and
> 3. how probability values are calculated and their significance, with reference to the individual case and in generally understandable terms.

[1] `http://www.bfdi.bund.de/EN/DataProtectionActs/Artikel/BDSG_idFv01092009.pdf?__blob=publicationFile` (accessed 07-09-2011)

> The first sentence shall apply accordingly if the body responsible for the decision
> 1. records the data used to calculate the probability values without reference to specific persons but creates such a reference when calculating the probability value, or
> 2. uses data recorded by another body.
> [. . .]

Furthermore, users must be able to change, add and delete their profile categories. This is also in line with what is stated in the Data Protection Directive 95/46/EC. For example, Google Dashboard, a tool for changing preferences, can only be used by users that have an Google account. Users that do not have an account can use the Google Ad Preference Manager to add/change/remove profile categories. Users can change how they are being targeted but they cannot stop Google from tracking them. For normal users it is hard to find where they can change their preferences because Google's Ad Preference Manager is not connected with, for example, the browser's privacy settings. Thus, data is already being collected before someone clicks on the turn off button in the preference manager from Google. The idea of letting users manage their preferences is commendable but a single place must be created where all settings and preferences for all kinds of advertising networks can be managed. Furthermore, it must be possible to get access to possible categories of interest that are deduced from the observed interests.

Very rich user profiles can be generated with group profiling. It is often the case that someone can be directly or indirectly identified and it is even more difficult to determine when an anonymous profile is becoming an identifiable profile. Therefore, group profiling should be covered by the new directive.

Nowadays, R&D departments come up with new technologies almost every day. Therefore, it seems necessary that the new directive has a broad scope and is technology independent. This idea is also supported by Mark Janssen (see section 3.6). Some tracking technologies are mentioned in the ePrivacy Directive 2002/58/EC but it is difficult to determine whether new technologies are covered by the directive. We suggest a technology independent solution. The new directive should cover all technologies that can be used for tracking in behavioral advertising. One advantage is that the new directive will not be outdated when advertising networks decide to use new technologies. Furthermore, it should not be allowed to use a tracking technology for any purpose before informed consent is achieved.

5.2 Privacy By Design

Privacy by design must be leading principle when we want to achieve what is stated in the directives. This section first deals with privacy by default.

Furthermore the idea of a new information standard for behavioral advertising is discussed and an example is given how it can be used. When we refer to the Article 29 Working Party we mean a letter[2] between the Working Party and the US Federal Trade Commission.

5.2.1 Privacy by Default

Privacy by default should be a leading principle for advertising networks, web site owners and browser vendors. Therefore, we will discuss a number of issues that have to be achieved before we can talk about privacy by default.

Advertising networks are the most powerful stakeholders because they are the data controllers that collect data and build information rich user profiles.

> User tracking must be turned off by default.

Therefore, if we want to achieve privacy by default tracking, and thus third-party tracking cookies, must be turned off by default. This can only be achieved if advertising networks and other organizations are forced by law not to track user by default. We explicitly state that user tracking must be turned off.

> Tracking technology independent solution.

By this we mean that all different kinds of tracking technologies, and thus data collection, are not allowed before a user has explicitly given consent. Thus, most notably deep packet inspection which will raise more and more privacy concerns in the future may not be used before a user has given consent. The Article 29 Working Party still discusses only third-party tracking cookies. In our opinion, there are lots of new tracking technologies that will replace cookies in the future.

> Only contextual advertising before a user has given consent.

However, contextual advertising can still be used to deliver relevant advertisements to the user. Many web sites, or parts thereof, deal with specific topics and could be served with advertisements that match with the sites content. A point of discussion is still how it can be achieved that public authorities or other supervising bodies can control if organizations comply with the principle privacy by default. We will not discuss this topic. In a nutshell, data collection may not be allowed for any kind of commercial or profiling purpose with regard to behavioral advertising before a user has given consent. How consent can be given will be discussed in section 5.2.3.

[2] `http://ec.europa.eu/justice/data-protection/article-29/documentation/other-document/files/2011/20110803_letter_to_oba_annexes.pdf` (accessed 03-09-2011)

5.2.2 Developing a Standard

It seems necessary to work on a standard that can be introduced for achieving informed consent. We will give some recommendations and ideas but we must also remark that a new standard will only be broadly accepted when law makers, advertising networks and browser vendors work together.

Using a standard format (e.g. XML).

We suggest that the new standard should be written, for example, in XML because it is a widely used language. The new standard should include all relevant information that has to be shown to the user. The new standard should include at least:

- a name and a corresponding address of the advertising network
- the purpose of data collection
- what kind of data is collected
- a link to the group profile(s) someone matches

Advertising networks must use the standard to inform users about their intentions. A link to a page where users can see the group profiles they match should be included. We have to mention that how a group profile should be presented to the user is a research topic on its own. We will not discuss this topic. Algorithms that are used to generate group profiles are a highly protected business secret. Research must still be done to get a better impression of what is possible.

All browser vendors should support the new standard.

As we already stated, a new standard can only be successful when it is broadly accepted. Therefore, the new standard should be supported by all browser vendors. It is thus a browser independent approach. Browsers will request the XML files of all corresponding advertising networks when a user visits a certain website. The information will be shown to the user and he/she can then decide whether to turn on/off a particular network.

Special folder for all XML documents.

Not only advertising networks have responsibilities but also web site hosts. Therefore, web site hosts should use a standard folder (URL) for all XML documents of the advertising networks they are in contract with. Storing the needed information for achieving informed consent at a central place makes it easier for browsers to find it and to show it to the users. This is also recognized by the Article 29 Working Party. Web sites use the main or index page as a starting point when someone visits the web site. The same could also be achieved for the information about the corresponding advertising networks. For example, `www.example.com/adinfo.xml` could include the destinations of all XML files of the different advertising networks on www.example.com.

5.2.3 What Browser Vendors should do

Browser vendors should cooperate with web site hosts and advertising networks.

> Implementation of the new standard.

Therefore, it is necessary that all browser vendors support the new standard in the future. We already mentioned in section 4.2.2 that different browser vendors developed their own PETs. For example, Firefox uses Do Not Track and Google Chrome uses ChromeBlock whereas both PETs have the same functionality. We do not want to force browser vendors to work on a solution that looks exactly the same between different browsers but we want to achieve that users are provided with sufficient information and that the information is easily accessible. Furthermore, it should be possible to import/export the preferences in case that users want to use another browser.

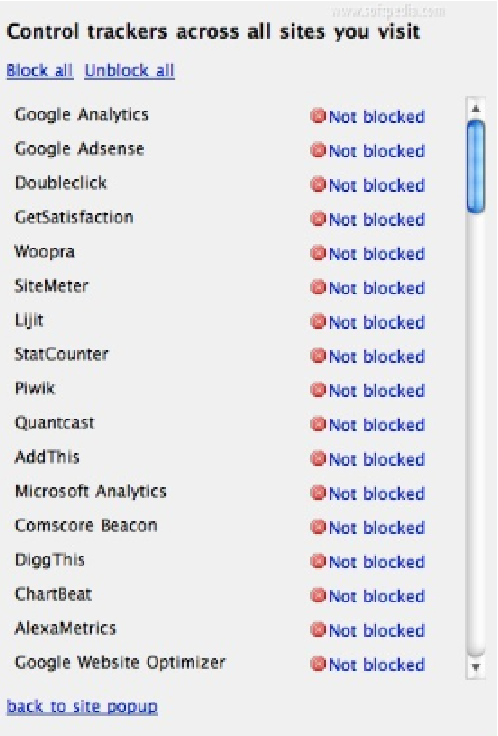
> Display advertising networks and trackers when a web site is visited

Informed consent can only be achieved when all advertising networks and trackers that are available on a web site are shown to the user.

In figure 5.1, Google's ChromeBlock lists all trackers of a web site in a popup. We support the idea of informing users with, for example, a popup. In figure 5.1, all trackers are enabled by default. This is not in accordance with European law. We support the idea of privacy by default and, thus, all trackers must be disabled by default. The browser requests the information of the different advertising networks from the web site and the popup is shown to the user for a few seconds when a web site is visited.

In our approach 5.2, the website provides the corresponding advertising networks to the user. ChromeBlock uses a list with advertising networks and examines to which connection is made.

> Consent can be given on a per advertising network basis.



Control trackers across all sites you visit

Google Analytics	Not blocked
Google Adsense	Not blocked
Doubleclick	Not blocked
GetSatisfaction	Not blocked
Woopra	Not blocked
SiteMeter	Not blocked
Lijit	Not blocked
StatCounter	Not blocked
Piwik	Not blocked
Quantcast	Not blocked
AddThis	Not blocked
Microsoft Analytics	Not blocked
Comscore Beacon	Not blocked
DiggThis	Not blocked
ChartBeat	Not blocked
AlexaMetrics	Not blocked
Google Website Optimizer	Not blocked

Figure 5.1: ChromeBlock[3]

We support the idea of the Article 29 Working Party that consent can be given on a per advertising network basis. It is not necessary to request consent for each access of information when the purpose of data collection remains the same. A link to the website of the advertising network where users can access the group profiles they match must be shown next to the name of the advertising network when a tracker is enabled.

Summarized, the advertising network gives the web site host the needed information in a XML document when a contract between the advertising network and the web site host is made. The web site host publishes all XML files on the web site. The browser accesses the XML files of the corresponding advertising networks when a web site is visited. All tracking technologies are disabled by default and the information about present advertising networks is shown to the user in a popup. The user decides whether to give an advertising network the right to collect data or not. The advertising network provides a link to the profile when consent is given.

[3] httphttp://i1-mac.softpedia-static.com/screenshots/ChromeBlock_1.jpg (accessed 03-09-2011)

5.3 Answering the Research Question

In this section, we will answer our research question. It seems that current legislation needs to be modernized because the average user does not know that he/she gets tracked almost every time when a website is loaded. Our approach (section 5.2) deals with making tracking more transparent for the user. Furthermore, we give recommendations (section 5.1) on what a reviewed data protection framework could look like. With regard to legislation, we give the following suggestions to achieve informed consent:

1. Modernize data protection framework

2. Public authorities must control compliance

3. Impact assessments

First, the data protection framework needs to be modernized. Informed consent can only be achieved when users are aware of that they are getting tracked and that they agree with it. Clear guidance must be given on how informed consent must be achieved. Furthermore, privacy by default should be a leading principle. Thus, our approach supports the idea of opt-in. Second, public authorities must be involved in the compliance process. They must have the right to control without prior notice if data controllers and data processors comply with the obligations. Third, impact assessments should be made when data controllers and data processors use the collected data for commercial purposes. Processing should be prohibited when most of the users are treated in a disadvantageous way. Also minorities, for example immigrants, may not be treated different.

Informed consent can only be achieved when the entire tracking and profiling process becomes more transparent. We suggest that all parties, advertising networks, web site owners and browser vendors, must have responsibilities. All have to contribute to process of making behavioral advertising transparent. We explicitly deal with making tracking, and thus behavioral advertising, more transparent. Users are getting the needed information (who wants to collect data, for which purpose and which kind of data) at a central place. The needed information is not longer hidden in privacy policies and terms of use. Finally, users can decide on the basis of the just gotten information if they want to allow a certain advertising network to track them.

We have to remark that the problem of getting profiling transparent remains. We will discuss this in detail in the next section.

5.4 Discussion

We already stated that behavioral advertising can be divided into two phases, a tracking phase and a profiling phase. Behavioral advertising will only be fully transparent for users when both phases are transparent as well. The profiling phase seems to be a well protected business secret. Advertising networks do not give users access to the group profiles they match. Profiling algorithms are kept behind locked doors because competitors could use them for their own advantage in case they are getting publicly accessible. It is not possible to examine the most likely non-distributive group profiles (section 2.2.2) which include suggestions about what a person could be interested in. Furthermore it is not possible to examine how detailed a group profile is. The degree of personal data that is included in a group profile is still an unanswered question. Also if there are any consequences for users when a group profile is applied, is still unknown.

All privacy critical tasks are done by the browser. The browser executes all decisions, for example block a certain advertising network, made by the user. Therefore, it has a high risk factor. When the browser is getting hacked, privacy settings could be changed and personal data sharing could be turned on.

First party advertising networks, like Amazon, are not directly covered by our approach (section 5.2) because they also use information (buying history of customers) for commercial purposes that is not collected by tracking a user. They can be included when they publish their own XML file with the corresponding information on their site. We have to remark that Amazon also uses the shopping history of a customer for behavioral advertising purposes. Our approach (section 5.2) will only work when this is illegal before a customer has given consent.

We give public authorities the power to conduct compliance audits. We do not give advice on how this can be done. Public authorities must have the right to access the databases of advertising networks. Whether the data structures and group profiles can be understood by auditors remains unanswered. We can assume that many different algorithms process data and therefore further research is necessary to examine how group profiles can be made transparent for both auditors and normal users.

Finally, advertising networks will only invest in new and expensive privacy-enhancing technologies when legislation requires it and is enforceable. As long as it is not clear what lawmakers are going to do in the future, advertising networks will remain conservative in spending money for better privacy.

5.5 Conclusion

Informed consent in behavioral advertising can only be achieved when advertising networks fulfill the obligations. Therefore, legislation has to provide clear guidance and address responsibilities. All parties should be incorporated in the legislation and design process which will maximize the level of acceptance. Furthermore, privacy by design and privacy by default as leading principles must be broadly introduced. Transparency plays an important role in the legislation and design process. Lawmakers have to provide a framework that protects the privacy of users and that is clear and comprehensive for all parties. Finally, the new framework has to be future-oriented because we already experienced in the past that new legislation can be outdated within months.

5.6 Future Work

We discussed profiling with regard to behavioral advertising but research on how group profiles are matched with individuals is still necessary. Also the question on how profiling can be made transparent for users is still unanswered.

We introduced a new information standard for behavioral advertising in section 5.2). Research is necessary on if and how this can be achieved technically. We give a framework about how the new information standard for behavioral advertising could look like. Research can be done on the development of the new information standard.

Furthermore, it seems interesting to implement our approach into a browser by means of a browser plugin. Our approach could be compared and evaluated with others like P3P.

Bibliography

[Ben01] C.J. Bennett, *Cookies, web bugs, webcams and cue cats: Patterns of surveillance on the world wide web*, Ethics and Information Technology **3** (2001), no. 3, 195–208.

[BvAA98] J.J.F.M. Borking, L. van Almelo, and M.J.T. Artz, *Gouden bergen van gegevens: over datawarehousing, datamining en privacy*, Registratiekamer, 1998.

[Byg01] L.A. Bygrave, *Automated Profiling: Minding the Machine: Article 15 of the EC Data Protection Directive and Automated Profiling*, Computer Law & Security Report **17** (2001), no. 1, 17–24.

[CK10] L. Curren and J. Kaye, *Revoking consent: A 'blind spot' in data protection law?*, Computer Law & Security Review **26** (2010), no. 3, 273–283.

[CLM$^+$02] L. Cranor, M. Langheinrich, M. Marchiori, M. Presler-Marshall, and J. Reagle, *The platform for privacy preferences 1.0 (P3P1. 0) specification*, W3C recommendation **16** (2002).

[CO03] L.J. Camp and C.A. Osorio, *Privacy-enhancing technologies for internet commerce*, Trust in the Network Economy (2003), 317–331.

[CO08] Jaquet Chiffelle and David Olivier, *Reply: Direct and Indirect Profiling in the Light of Virtual Persons. To: Defining Profiling: A New Type of Knowledge?". In Hildebrandt, Mireille; Gutwirth, Serge. Profiling the European Citizen*, Springer Netherlands, 2008.

[Com10] European Commission, *Communication from the Commission to the European Parliament, the Council, the economic and social Committee and the Committee of the regions*, November 2010, COM(2010) 609 final, available at: `http://ec.europa.eu/justice/news/consulting_public/0006/com_2010_609_en.pdf`.

[Con08] Greg Conti, *Googling Security: How Much Does Google Know About You?*, 1 ed., Addison-Wesley Professional, 2008.

[Cus04] B.H.M. Custers, *The power of knowledge; Ethical, Legal, and Technological Aspects of Data Mining and Group Profiling in Epidemiology*, Wolf Legal Publishers, 2004.

[Deb05] F. Debusseré, *EU E-Privacy Directive: A Monstrous Attempt to Starve the Cookie Monster*, Int'l JL & Info. Tech. **13** (2005), 70.

[DF03] J.E. Dobson and P.F. Fisher, *Geoslavery*, Technology and Society Magazine, IEEE **22** (2003), no. 1, 47–52.

[Dwy09] C. Dwyer, *Behavioral targeting: A case study of consumer tracking on levis.com*, Pace University (2009).

[Eck10] P. Eckersley, *How unique is your web browser?*, Privacy Enhancing Technologies, Springer, 2010, pp. 1–18.

[FTC10] FTC, *Protecting Consumer Privacy in an Era of Rapid Change*, Tech. report, Federal Trade Comission, December 2010.

[Gol] Ian Goldberg, *Privacy Enhancing Technologies for the Internet III: Ten Years Later.*

[GT11] A. Goldfarb and C.E. Tucker, *Privacy regulation and online advertising*, Management Science **57** (2011), no. 1, 57.

[Hil08] M. Hildebrandt, *Defining profiling: a new type of knowledge?*, Profiling the European Citizen (2008), 17–45, Springer Netherlands.

[HN00] D.L. Hoffman and T.P. Novak, *Advertising pricing models for the World Wide Web*, Internet publishing and beyond: The economics of digital information and intellectual property (2000), Cambridge, UK: MIT Press.

[HN09] D.C. Howe and H. Nissenbaum, *TrackMeNot: Resisting surveillance in web search*, Lessons from the Identity Trail: Anonymity, Privacy, and Identity in a Networked Society (2009).

[KtH11] R.F. Koorn and J. ter Hart, *Privacy by Design: From privacy policy to privacy-enhancing technologies*, Compact IT Advisory (2011).

[Kun07] C. Kuner, *European Data Protection Law. Corporate Compliance and Regulation*, Oxford University Press, 2007.

[Kun08] ———, *The 'Internal Morality' of European Data Protection Law*, November 2008.

[MNS11] J. Mayer, A. Narayanan, and S. Stamm, *Do Not Track: A Universal Third-Party Web Tracking Opt Out*, Tech. report, Network Working Group, March 2011.

[MW10] B. Masiello and A. Whitten, *Engineering Privacy in an Age of Information Abundance*, 2010 AAAI Spring Symposium Series, 2010.

[NF09] Andrew Newman and Miriam Fonfe, *Booking Targeted Ads Efficiently: Managing reach and frequency to maximize response rates in online advertising*, Yahoo Insights Snapshot, June 2009.

[oE10] Council of Europe, *Draft Recommendation on the Protection of Individuals with regard to Automatic Processing of Personal Data in the Context of Profiling*, June 2010, 26. Plenary Meeting, Strasbourg.

[Para] Article 29 Data Protection Working Party, *Opinion 1/2010 on the concepts of "controler" and "processor"*, `http://www.cbpweb.nl/downloads_med/med20100219_C.03%20DC-DP_Opinion_ADOPTED.pdf`.

[Parb] ______, *Opinion 13/2011 on Geolocation services on smart mobile devices*, `http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2011/wp185_en.pdf`.

[Parc] ______, *Opinion 2/2010 on online behavioral advertising*, `http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2010/wp171_en.pdf`.

[Pard] ______, *Opinion 4/2007 on the concept of personal data*, `http://ec.europa.eu/justice/policies/privacy/docs/wpdocs/2007/wp136_en.pdf`.

[Roo] A. Roosendaal, *Facebook Tracks and Traces Everyone: Like This!*, Tilburg Law School Research Paper No. 03/2011.

[SHKV08] Wim Schreurs, Mireille Hildebrandt, Els Kindt, and Michael Vanfleteren, *Cogitas, Ergo Sum. The Role of Data Protection Law and Non-discrimination Law in Group Profiling in the Private Sector*, Profiling the European Citizen (Mireille Hildebrandt and Serge Gutwirth, eds.), Springer Netherlands, 2008, 10.1007/978-1-4020-6914-7_13, pp. 241–270.

[SNE] S. Steiniger, M. Neun, and A. Edwardes, *Foundations of location based services*, CartouCHe Lecture Notes on LBS, version **1**.

[Sot08] S. Sottiaux, *Terrorism and the limitation of rights: the echr and the us constitution*, Human rights law in perspective, Hart, 2008.

[TI11] TNO and IViR, *A bite too big: Dilemma's bij de implementatie van de cookiewet in Nederland*, TNO-rapport 35473, February 2011, Available at: `http://www.tno.nl/content.cfm?context=overtno&content=nieuwsbericht&laag1=37&laag2=2&item_id=2011-03-15%2014:36:08.0`.

[TKH+09] Joseph Turow, Jennifer King, Chris Jay Hoofnagle, Amy Bleakley, and Michael Hennessy, *Americans Reject Tailored Advertising and Three Activities that Enable It*, September 2009, Available at SSRN: `http://ssrn.com/abstract=1478214`.

[TvE] H. Tschofenig and R. van Eijk, *DO NOT TRACK. An Attempt to Frame the Debate*.

[vBBO03] G.W. van Blarkom, J.J. Borking, and J.G.E. Olk, *Handbook of Privacy and Privacy-Enhancing Technologies*, Privacy Incorporated Software Agent (PISA) Consortium, The Hague (2003).

[vEZ04] Minister van Economische Zaken, *Besluit Universele Dienstverlening en Eindgebruikersbelangen*, May 2004, WJZ 4028595.

[WH09] FIDIS WP7 and M. Hildebrandt, *D 7.12: Behavioural Biometric Profiling and Transparency Enhancing Tools*, March 2009, Available at: `http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp7-del7.12_behavioural-biometric_profiling_and_transparency_enhancing_tools.pdf`.

[WHM06] FIDIS WP7, M. Hildebrandt, and M. Meints, *D 7.7: RFID, Profiling, and AmI*, August 2006, Available at: `http://www.fidis.net/fileadmin/fidis/deliverables/fidis-wp7-del7.7.RFID_Profiling_AMI.pdf`.

Applicable Law

[ECHR]	European Convention on Human Rights
[CFREU]	Charter of Fundamental Rights of the European Union
[95/46/EC]	Data Protection Directive 95/46/EC
[2002/58/EC]	ePrivacy Directive 2002/58/EC
[2009/136/EC]	Directive 2009/136/EC of 25 November 2009 amending Directive 2002/22/EC on universal service and users' rights relating to electronic communications networks and services, Directive 2002/58/EC concerning the processing of personal data and the protection of privacy in the electronic communications sector and Regulation (EC) No 2006/2004 on cooperation between national authorities responsible for the enforcement of consumer protection laws