What Deters Jane from Preventing Identification and Tracking on the Web?

Fatemeh Shirazi  
TU Darmstadt/ CASED  
fshirazi@cdc.informatik.tu-darmstadt.de

Melanie Volkamer  
TU Darmstadt  
melanie.volkamer@casede.de

ABSTRACT

Recent revelations about surveillance by several institutions and de-identification can be expected to have increased public awareness of identification- and tracking-related privacy threats. It is reasonable to expect that the general public has started using corresponding privacy protection mechanisms. Our goal with this research was to determine whether they actually do this. If not, we wanted to explore possible explanations for not uptake such privacy-protecting countermeasures. We interviewed 20 (mainly lay) people and found that our interviewees did not proactively try to prevent being identified and tracked. We identified seven different types of explanations. Including a number of misconceptions which might explain this puzzling level of apathy. The participants demonstrated confusion between different kinds of sensitive data; and displayed a confusion between the semantics of ‘privacy’ and ‘security’. The findings also indicate that security being compromised, resulting in losing money for example, is more concrete and more easily brought to mind than privacy-related problems. In terms of the consequences of surveillance, the most commonly cited outcome is the receipt of personalized advertisements, which many consider beneficial. Potentially negative impacts of identification and tracking is often assumed to not occur to them. Our interviews also pointed out a gap between passive and active knowledge about identification and tracking techniques, their impact on privacy and countermeasures against them.

Categories and Subject Descriptors

K.4.1 [Computers and Society]: Public Policy Issues

General Terms

Privacy, Human Factors

Keywords

Privacy, Identification, Tracking, Mental Models, Concerns, User Behaviour Analysis

1. INTRODUCTION

Various privacy problems exist on the Internet. Examples are non-consented data collection (which includes identification and tracking on the web), re-purposing data and profiling (often based on unprotected content, for example from social networks), and data persistence (even after an active deletion from the data owner). The focus of our research is identification and tracking throughout network activities (i.e. without being logged in).

Over the last few years researchers have developed a variety of tools enabling privacy protection against this kind of privacy threats. Examples are proxies\(^1\), Tor\(^2\), and various types of browser add-ons\(^3\). Web browsers routinely also allow users to disable or delete cookies or to browse privately\(^4\), all of which prevent identification and tracking or at least make it harder.

The wide-scale surveillance of ordinary citizens by foreign and local governments has been revealed by the press over the last few months\(^5\). Price discrimination techniques based on identification and tracking were also discussed in the press\(^2\). With this level of media visibility one could reasonably expect Internet users to be more aware of identification and tracking related privacy threats. Awareness, should, we anticipate, lead to increased deployment of widely available privacy protection mechanisms in order to prevent or deter such privacy violations.

Is this actually the case - are people aware and do they protect their privacy against identification and tracking on the web? If not, why do people not protect their privacy? Past research refers to usability issues of many of the available tools\(^3\),\(^4\) and technological reasons such as a decrease in efficiency or less functionality\(^2\). This reflects a common misconception, that non-use is primarily due to non-usability. For years there has been a widely-held assumption that when security and privacy protecting tools

1. www.anonymizer.co/  
2. www.torproject.org  
3. getadblock.com, adblockplus.org  
are usable and efficient people will start using them. A recent publication suggested that this view was naive in the context of end to end email encryption, that a number of other reasons for non-uptake were identified. We wanted to gain a greater understanding of the reasons people have for not using privacy-protection tools. With this understanding we can develop strategies to encourage use thereof.

Therefore, our research goal is to derive a more complete picture of possible explanations for not using available privacy protecting tools. Our hypothesis was that non-use is in addition due to misconceptions and a lack of awareness. Therefore, we also contemplated those communication channels that are most likely to reach people to increase awareness and correct misconceptions. Finally, we arrive at recommendations for guiding future work to ameliorate the current situation.

1.1 Methodology
We applied an exploratory approach and conducted semi-structured interviews in Germany with 20 people (mainly lay people but also two experts). Questions regarding known threats, type of entities trying to violate privacy and why, own consequences, available and used countermeasures, issues they see with available tools, and communication channels to reach people and inform them. The interviews were transcribed and mentioned themes (in terms of explanations or communication channels) were identified independently by the authors and discussed in order to agree on those reported in this paper.

1.2 Result
In total, seven different explanations, some with a number of sub-themes, were identified: (1) People primarily correlate privacy issues with issues not related to identification and tracking (and thus corresponding countermeasures do not help to protect against identification and tracking on the Internet); (2) People are not aware that meta information is actually sent with each webpage request; (3) People who are aware of meta data being transmitted tend not to be aware that such information can be used to identify and/or track them (or how); (4) People who are aware that identification and tracking is possible using their meta data are not concerned for various reasons. The first is because they feel that they are not important enough or have nothing to hide. The second is that they are not aware of who and why different entities would want to identify and track them other than to produce personalized advertisements. Third, they are not aware of actual consequences (other than personalized advertisements); (5) People are not aware of existing effective countermeasures; (6) People are not able to use them properly; (7) People are able to use them, but become side tracked for other reasons.

We also found that people did indeed want to know more about potential privacy threats, their consequences and how they ought to protect themselves.

2. RELATED WORK
Related work often focuses on one of two areas: usability and mental models in terms of awareness and possible misconcepts. Research in the former area identifies usability and technical issues that keep users from deploying different privacy protecting or security tools. The other potential causative is users’ general understanding of privacy or security threats (so called mental models) in different contexts, which helps to explain why specific tools are not as widely used as anticipated. There are only few papers available trying to draw an exhaustive picture of possible explanations; e.g. in the context of end-to-end email encryption [25]. Usability evaluations have been conducted in the context of privacy-protecting tools to protect against online behavioral advertising [14] and the effectiveness of tools limiting online behavioural advertisements [5]; in the context of the anonymous communication tool, Tor [21]. Usability and technical issues in the context of security tools has been studied in a number of different applications: e.g. in the context of end to end email encryption usability issues have been identified in [34] and [7] and interoperability issues in [20]; in the context of firewall configuration usability [35]. Acquisti et al. have investigated in the inconsistencies between privacy decisions making and the actual behaviors of consumers focusing on privacy issues raised in e-commerce [3]. In their work the point at several reasons for these inconsistencies among them are lack of sufficient knowledge and also that consumers often “are likely to trade off long-term privacy for short-term benefits.”

All of them report flawed end-user mental models, leading to an incomplete or flawed understanding of threats, consequences and/or effective general countermeasures. Therefore, it is important to help people to construct correct mental models of privacy in order to facilitate privacy decisions as also stated by Coopamootoo et al. [7].

3. BACKGROUND
Solove categorized privacy threatening technologies into non-consented data collection, data processing, and data dissemination [27]. The techniques we review fall all under data collection. In particular, how users are identified and how their online activities are (implicitly) tracked. Tracking technologies and countermeasures against them have been extensively reviewed in the literature [17, 13, 8]. In this section we review commonly used identifying and tracking techniques (possible consequences for tracked people) and their countermeasures.

Readers being familiar with both techniques and countermeasures can skip this section.

3.1 Techniques to identify and track users
We classify the identifying and tracking techniques into two main categorizes according to [18]: statefull and stateless tracking. In statefull techniques the users’ device are usually actively involved and such techniques are often used by web page providers. For stateless techniques the users’ device has no active involvement. Therefore, the later techniques can be applied by web page providers as well as other entities such as any network adversary or Internet Service Providers. The following stateless and statefull techniques (and corresponding consequences) exist:
IP addresses: The Internet Protocol (IP) address is a label assigned to each device that is connected via the Internet Protocol. IP addresses are the most common user identifiers on the web. They are used for all interactions between the user’s device and the web. However, IP addresses cannot always be considered a reliable identifier.

Browser fingerprinting: Another technique to identify people on the web is by the set of device attributes of the user’s browser, which built the so-called browser fingerprint. Based on the browser configuration such as fonts, screen resolution, language, etc., a client’s browser can often be identified within manageable group of browsers [12, 2].

Cookies (http cookies, flash cookies): For data communication on the web, HTTP (HyperText Transfer Protocol) protocol, is a stateless and operates on a request-response basis. Hence, HTTP requests are treated independently from each other. Therefore, two HTTP requests which have been sent by the same user could not be correlated. In order to keep track of the sender of an HTTP request in a browser-server communication, cookies have been introduced as an addition to HTTP. A cookie is a text-only string that is entered into the browser’s memory; the server and the user’s device pass the cookie back and forth during their interactions in order to keep track of the state of their interactions. HTTP cookies can be also set by third parties. Another type of cookies are Adobe Flash Local Shared Object, Flash cookies, where information is stored on a user’s computer to facilitate Adobe Flash applications, i.e. user preferences. The local shared objects have also the same functionality as HTTP cookies, containing attributes for unique identifiers [29, 16, 29]. Flash cookies are stored outside the browser’s memory. Therefore, the browser has little control over the Flash cookie. Hence, by deleting the cookies on the browser’s setting, the Flash cookies are not affected.

Web bugs and JavaScript/Page Tags: Web bugs (also called web beacons) are used to inform the web bug creator when a web page is visited or an email is opened. Web bugs are small images (usually transparent, 1 * 1 pixel) embedded in a web page. The image is handled like any external image in a web page; external images (also ads, videos, etc.) are requested from the source web server. When a web page or email (uploaded from a so-called first party server) contains a web bug, the browser sends an HTTP request to the web bug source server (third party server) in order to retrieve the image. A JavaScript tag is a JavaScript code which is embedded into a web page. When the web page is loading, the JavaScript code is downloaded and then executed. The JavaScript code can be programmed to send user data to the data collection server. Page tags can be used to collect and send the history of visited links (so-called history sniffing), click data, page view, and cookies. They also can be programmed to update cookies.

Consequences. Privacy related consequences have been reviewed in the literature from several aspects. Some examples are as follows: Odlyzko discusses the advantage and disadvantages of online price discrimination [22] and mentions insurances also as example. Vissers et al. investigated in price discrimination for airline tickets [31]. Another consequence is targeted advertisements (also from third parties). and anotheSolove points the use of personal data in political campaigns and researchers found out that web pages associated with governmental offices perform very persistent tracking.

3.2 Countermeasures

In this section we review basic and commonly used countermeasure actions that users can perform to handle the above mentioned techniques.

Hiding IP address: users may hide her IP address from the web services by using a simple proxy or anonymous communication networks. A proxy server is a server acting as an intermediate when a user is requesting a resource like a web page. Anonymous communication networks extend the idea behind proxies; consisting of a set of routers, often selected randomly, which relay the communication until it reaches its destination. Using anonymous communication networks users can conceal the destination of their communications towards local adversaries, e.g., their ISP, as well as protect their identity towards the destination itself [10]. Examples of anonymous communication networks are Tor [11] and I2P [1].

Protecting against browser fingerprinting: There has been few countermeasures proposed for protection against browser fingerprinting. Acar et al. have investigated in using the Tor browser, which allows only restricted and Fire-gloves, a research-purpose browser extension for Firefox that sends “randomized values when queried for attributes” [2].

Managing cookies: the user can manage cookies in the browser setting directly, or by using cookie managing tools, she may use certain tools that identify tracking cookies such as “track-your-tracker” 8, using browser privacy modes where locally the browsing history is disabled, and local traces are reduced (Firefox: private browsing, Google Chrome: Incognito mode, IE:InPrivate mode). In addition, browser extension can be used such as AdBlock Plus 9 which is a add-ons which disrupts the interaction between the user’S device and advertisement servers.

Protecting against web bugs and JavaScript tags: In order to counteract against such techniques people may use privacy supporting browser extensions that offer selective blocking of corresponding client-side content, i.e. JavaScript, Flash, and web bugs. Some famous examples of such extension are: NoScript 10 which enables selective blocking of contents such as JavaScript and Flash. Ghostery 11 identifies cookies, scripts, and web bugs and offers the option to block them [14].

4. METHODOLOGY

We performed an exploratory study consisting of semi-structured interviews, and subsequent qualitative analysis of corresponding transcripts (of audio data) in order to evaluate: (first part of the research goal) which privacy protecting tools are in use; and which not; (second part of the research goal) what are possible explanations of not using such tools; (third part of the research goal) what would be effective communication channels to inform users about threats

---


8www.sit.fraunhofer.de/de/track-your-tracker

9adblockplus.org

10noscript.net

11www.ghostery.com/de
and countermeasures. The interviews were conducted in German in early summer 2014 and questions and answers translated for this paper. A number of pre-interviews were conducted to improve the interview guidelines iteratively.

4.1 Study Protocol
Participants went through the following phases:

**Welcome and general instructions.** Subjects were welcomed and the goal (privacy issues on the Internet due to implicit data collection and processing / caused by identification and tracking). The exploratory nature of the research was explained. They were asked to think aloud and told that there were no wrong answers to the questions. They signed the consent form (which, if agreed, also allowed us to record the interview).

**Exploring general level of expertise.** In the first part, subjects were asked to describe what happens on the Internet as well as which information is sent when they click on a link in an email. First a http/Amazon and second a https/Buch24 (popular online book shop) link. The answers helped us to assess their level of expertise with respect to Internet technologies.

**Tracking mental model.** In this first main part, subjects were asked questions related to their mental model on tracking-related privacy issues. We asked about known threats and, for each mentioned threat, we asked who caused the mentioned threats, what their motivation for these were and which countermeasures were known.

**Knowing versus using privacy protecting tools.** Afterwards, we asked the subjects for each of the mentioned countermeasures whether or not they actually used them and why (or why not).

**Concerns.** After having made them think about threats and what they actually do to counter these, we asked them to state how concerned they were on a three point scale.

**Tracking examples.** In this part, we discussed concrete scenarios with the subjects. First we showed them a German-wide weather forecast page that displayed the weather forecast for the city in which we conducted the interviews (facilitated by the IP address). Second, the amazon page recommending items based on a previous search without being logged in (facilitated by a cookie). Third, the MSN page, also showing items based on an earlier search. For each scenario, we asked them to come up with an explanation for the tailoring, and for them to articulate their concerns. (because of third party cookies and web bugs).

**Threat knowledge quiz.** Afterwards, we mentioned threats that they hadn’t mentioned up to that point. For each, we asked whether they were aware of the threat and, if yes, whether they could explain it. Threats that were mentioned are: being tracked by IP address, browser fingerprinting, cookies (third party, Flash cookies), web bugs, and JavaScript tags. This section was included to avoid missing threats that people were aware of, but did not mention earlier.

**Browser privacy settings.** The focus of this part was the browser’s privacy settings as first countermeasure. Subjects were asked whether they had seen the dialog before (depending on the browser they were used to, we showed them a screenshot of Firefox, Chrome or Internet Explorer privacy setting dialog), and whether they had made some changes to their browser (and if so, why).

**Browser privacy mode.** For the same reason, we addressed the web browser’s privacy mode. Subjects were asked whether they were aware of this mode and whether (and if so in which situations) they used it (or would use it). We also asked subjects to read the explanations provided by the different browsers for this mode and asked whether they noticed the differences with and without privacy mode for Firefox, Internet Explorer and Chrome.

**Knowledge quiz - countermeasures.** Afterwards, additional countermeasures were mentioned. In case subjects stated that they are aware they, were asked to explain them and asked whether they use them (including why or why not).

**Concerns.** At this stage, having been primed with the discussion of threats and countermeasures, participants were asked whether they were now more or less concerned than before.

**Communication Channels.** Participants were asked how they would inform themselves about privacy threats and countermeasures and what they thought were effective communication channels.

**Demographics.** We asked subjects for their age and their type of job/course of study.

After the official part (no longer recorded), participants could ask questions and received some helpful explanations about privacy threats and countermeasures.

4.2 Recruiting and reimbursement
Our goal was to interview broad variety of people with different backgrounds, age ranges, and level of knowledge and expertise in the context of privacy and tracking in particular. We asked our friends to forward a corresponding invitation email / Facebook post to other non-computing science friends whom they thought would be interested in participating, using a snowballing technique. We asked them not to tell their friends about the type of research we were conducting. Interested people were asked to contact us for an appointment. In selecting 18 people, we tried to balance gender, background, and whether they are students or not. We also included two computing experts in the interview group. We selected two who were not part of our research group but in the CASED research center. The interview took between 30 and 40 minutes. The shortest interview lasted 24 min, the longest 75 min. On average, an interview took 40 min. Participants were reimbursed with 10 Euro.

4.3 Ethical considerations
Ethical requirements for research involving human participants are provided by an ethics commission at the university.12 The relevant ethical requirements regarding participant consent and data privacy were met. Participants were first informed about the purpose of the study, after which they could decide whether or not to proceed with the interview. Also they were told that they could stop the interview at any time. In order to meet the data privacy requirement, a privacy statement was provided stating that participants’ data would only be collected for research purposes, their identity would not be linked to their responses, and their data would not be passed on to third parties. Furthermore, participants’ data was only handled by members of our research group.

4.4 Evaluation methodology
All participants agreed to recording of the interview and the tapes were transcribed — some be the authors, some by other members of the research group. Responses were analyzed using open coding. Open coding (instead of closed coding) was used in order to obtain emerging themes from participants’ responses, rather than beginning the data analysis with pre-selected themes due to the exploratory characteristic of our research. If one participant mentioned several concepts, each was coded under an appropriate theme.

Both authors independently reviewed a subset of the transcripts and identified themes (categories of possible explanations) and sub-themes (individual explanation) from participants’ responses. The themes and sub-themes were discussed and iteratively developed. Finally, we agreed on the six themes and several sub-themes reported on in following subsections. The remaining responses were then analyzed and coding to sub-themes compared. Since we are reporting qualitative research results, like Wash [32], we will not report how many users alluded to each sub-theme. We do attempt to give a flavour by using terms like ‘very few’ (1-2 participants), ‘some’ (3-10), ‘many’ (11-18), and ‘most’ (19 or 20). We substantiate the various themes and sub-themes with (translated) quotes where applicable.

4.5 Interview group
The interview group consisted of 20 participants: eleven female and nine male. About one third were students and two privacy researchers from our research institute. The remaining had different kind of jobs. The students studied various subjects with four studying psychology (or psychology in IT) and three education science. All other subjects appeared only once. Eight participants were in the age group 19-23, eight in the age group 24-28 and the remaining four were between 29 and 60 years of age.

Note, due to the exploratory characteristics, we did not seek participants representative of the German population. As such, we are able to present a set of explanations from the selected interview group while we do not claim that the list is complete. It might be representative for young people but it would be necessary to interview more older people to evaluate whether further explanations can be identified.

5. RESULTS
The following seven themes i.e., possible explanations, and a number of sub-themes were identified.

5.1 Privacy issues do not correlate with identification and tracking issues
The first theme or explanation that was identified is: People’s privacy issues do not correlate with identification and tracking issues on the Internet and thus their corresponding countermeasures do not help to protect against identification and tracking on the Internet. Applying any countermeasure always requires some effort. Thus, people primarily take action against those threats they can immediately think about. We identified the following sub-themes with threats that are more prominent to them:

Only Explicitly provided data is at risk: Most participants primarily thought of explicit data, either pictures etc provided in social networks or bank account details (or indirectly buying habits), when considering privacy threats. ‘the worries are rather concerning Facebook and because one doesn’t protect oneself there’ (16). Correspondingly, their applied and/or proposed countermeasures were either careful usage of social networks and/or selection of online shops or not using such services at all. Other mentioned countermeasures were to use secure passwords.

Tracking is only possible while logged in: Some people think that tracking on the Internet is only possible while they are logged in by those they have accounts with. ‘If one doesn’t log in yourself then you stay anonymous’ (11). Furthermore, they believed that the different services they had accounts with collaborate and interchange data with each other; e.g. in order to display personalized advertisements: ‘Google and Facebook and Twitter, which put all their data together’ (1). Correspondingly, they try to remember to log out whenever they have finished using a service.

Privacy violations by malware: Some people have mainly concerns that viruses, Trojan horses and other type of malware is running on their devices and reading the entire network communication. ‘I know that one can download a Trojan or virus, which saves every thing’ (9). Correspondingly they protect their privacy by using anti virus scanners and carefully using the Internet. Very few think that bought software also collect data to recommend further products. Those recommend to use open source software if available.

Misconceptions about potential privacy violations: Some people mentioned privacy threats which are actually primarily not privacy threats, such as Phishing. When asked why this was a privacy threat, most participants acknowledged that it is actually not a threat. Only one of them mentioned that Phishing was about getting money out of people while privacy was about tracking purchases. ‘probably they only want my money, though they would be able to deduce what I have bought on PayPal or Amazon.’ (2).

Summary: The reported sub-themes show that one reason for not protecting against identification and tracking might be that people’s are not primarily concerned about their implicit data but more concerned about their explicit data, in particular after having logged in. Note, our general impression is that people have heard of threats and countermeasures but more in terms that they came across the terms (e.g. ‘Flashcookies, yes I have heard, but what it is, no’ (8)). They don’t seem to have any clear picture of privacy threats, who invades their privacy or how to protect themselves.

5.2 People are not aware what information is transmitted
The next theme or explanation that was identified was that people are not aware of (meta) information which is actually transmitted with each web page request and thus should be protected or blocked. Knowing what is sent is essential in order to know what actually needs to be prevented.

Only the expert participants were aware of web bugs and java script tags while the remaining participants were not (although once we mention java script tags they assume it is related to java script, that a few did block). Only one of the experts mentioned more advanced information that might be sent to the web service, like an identifier to the source of a link (i.e. from which email or on which webpage one clicked on the link).

Many participants were not aware of browser settings being transmitted with each request. Those non-experts who were assigned to this sub-theme were talking about ‘special
5.3 People are not aware how information is used to identify/track them

The third theme or explanation that was identified is: People who are aware of sent meta data are not aware how various types of (meta) information can be used to identify and/or track them and thus should be protected. It is important to know how each of the (meta) information can be used to violate privacy - both on the Internet and by the actual web service provide (without being logged in). We report the identified sub-themes for the different type of meta information separately.

5.3.1 IP address

Only very few really know what an IP address is and that it is mainly used to be tracked or to generate profiles as stored together with the requested content. We identified the following two sub-themes which might serve as explanation for not effectively protecting IP address information:

Long-range location information that is not privacy critical: Some participants believe that IP addresses only provides information of the country, region or city 'similar to the ZIP code' (15); 'country and region' (13). As such it is not considered as privacy critical information but information used to improve the service (like in the weather forecast example).

Computer’s identification number that is privacy critical and thus should be deleted from time to time: Some participants believe that IP addresses are more powerful as they actually are as they believe that it is a unique number for a PC and is used by web service providers (like Amazon) to identify or track people: e.g. ‘But my computer has a specific identifier. Meaning when I go from my sisters computer on Amazon, without logging in, something like this does not show’. As a consequence a few participants do delete or recommend to delete IP addresses from time to time.

Identification number that is not privacy critical: Some participants being aware of the sensitivity of IP addresses are not concerned as they believe that it can only be linked to a person if necessary to solve a crime: ‘who I am personally is just possible through court order’ (9).

Summary: The reported results show that one reason for not protecting IP addresses might be the lack of awareness about how it can be used to identify and/or track people.

5.3.2 Cookies

Very few really know what the different types of cookies did or how they are used. We identified the following two sub-themes which might serve as explanation for not effectively protecting IP address information:

Cookies cause delays: Some have no idea what the purpose of Cookies are but some learned that they should be deleted from time to time to increase the computer’s efficiency. ‘It slows the PC down’ (1); ‘I have often deleted cookies because I got annoyed by popups and I didn’t know if cookies were causing that’ (15); ‘These things make the computer slow too’ (1). Thus, these people only delete cookies if they feel that their computer is slowing down.

Cookies cache data: Some think cookies are the type of information stored on the computer that is often used (i.e. data is cached on the computer) and thus does not need to be downloaded again and again. ‘downloaded stuff are saved in them, so that on the next time they don’t have to be downloaded again from the server but can be retrieved from my computer’ (8). Thus, there is no need to remove cookies.

Cookies store the browser history but are mainly used to personalise advertisements: From some participants’ statements, we got the impression that cookies are equal or similar to the browser history; that this information is stored locally and the entire history is sent to the service providers. While these people should see a need to protect cookies or block them from being sent, they are often not concerned about their privacy because they think service providers only collect this information to provide personalized advertisements. Some also think google tracks them to make search requests more efficient (see Section 5.4 for findings why people are not concerned about this). From others, we got the impression that they believed that without cookies they would also disable the history. They consider the history functionality to be useful.

Third Party Cookies are not understood but annoy people: Most people are aware that they see personalized advertisements on third party websites while most of them did not understand how this happened (which date is shared, interchanged by whom). Some stated that they used an ad-blocker to avoid these types of advertisements, but mainly because they were annoyed and not because they are concerned about their privacy. ‘I assume, they block advertisements’ (18); ‘because I just dont want to see advertisements’ (19).

Flash Cookies are related to the FlashPlayer: While only the experts mentioned Flash cookies, some people stated that there is most likely a relation to the (activation of) FlashPlayer software: ‘they are probably cookies which are related to FlashPlayer’ (10); ‘That they enter code through the Flashplayer’ (4).

Summary: While the term ‘Cookie’ is known, only very few fully understood what cookies do and how they are used to track them. This confirms the findings of McDonald et
al. [19] in regard to users’ understanding of cookies. Correspondingly, it is not surprising that people did not block or delete cookies other than to prevent being annoyed by third parties showing personalized advertisements.

5.3.3 Web bugs, java script tags, etc

Only very few are aware of what advanced techniques are used and how. Sometimes a correlation with java script being active or not but without knowing what the problem is as they were only aware of malicious software being installed if java script is enabled. Once explained - surprisingly - people were concerned by web bugs in terms of that there are transparent pictures that they cannot see: ‘this with the invisible image, shocked me a bit, because it is invisible and how am I supposed to recognize that, if I dont know that it exists?’(1).

Summary: Obviously, people do not protect against these advanced techniques if they are not aware that they exist.

5.4 People are not concerned of being identified and/or tracked

We identified in total five sub-themes that might help to explain why people to not see a need to protect their privacy against identification and tracking on the Internet.

...because they think not to be important enough: Some participants are not concerned because they think they are not important enough in particular in relation to the required effort to protect their privacy: e.g. ‘small unimportant account... I am not important enough...taking effort for what I use the Internet for’ (2), ‘being an average citizen’ (12) Very few of them also mentioned that they know that this is not an appropriate attitude.

...because they think they have nothing to hide: Some participants stated that they are not concerned as they ‘do not have real secrets’(7), ‘sent data are not considered as bad’(10), and as they do ‘nothing illegal’ (12)

...because they care more about financial data: We noticed that some people are more concerned about their financial data than other types of sensitive information: ‘The only thing that I am worried about is Internet-banking, when it is about money. Here I am really worried.’(7).

...because they are not aware of who identifies and tracks for which purpose: Many participants consider personalized advertisement as the only reason why service providers want to identify and track them. Some also consider governmental agency (‘states’, ‘NSA’, ‘BND’). Internet service providers are only mentioned by very few and only in the context of them providing information to governmental agencies in case of a crime. This number was for us surprisingly low as there have been several debates in the news about the so called ‘Vorratsdatenspeicherung’. Besides crime the other mentioned but not further elaborated reasons for governmental agencies to identify or track people is for espionage reasons. However, they are not concerned about this mainly because ‘it does not effect me’ (8) or because they cannot protect against them (see also Section 5.5). E.g. one participant stated ‘So I cannot imagine that tomorrow my house bell is ringing, and the BND stand there and takes me with them. I do not see that for myself’(7).

... because they are not aware of the consequences: Many participants seem not be aware of the consequences of identification and tracking on the web. Some stated this explicitly: e.g. ‘Because I cannot imagine what kind of consequences it can have.’(16); ‘What does someone want to do with my data. I really cannot imagine that’(7).

Summary: The last two sub-themes somehow encompass the rest: People are not aware of potential consequences. This lack of awareness leads to other explanations such as not being important enough or not having anything to hide as also explained in [28]. Furthermore, we gained the impression that actually most people stated that they were unconcerned but could not explain their lack of concern. It is more of a general feeling than being afraid because of concrete negative consequences. Moreover, some people asked after the interview, off the record, for more information about possible consequences, which indicates a general lack of awareness.

5.5 People are not aware of countermeasures

Before reporting participants’ awareness of concrete effective countermeasures, we report the sub-themes we identified in the general context of countermeasures.

There is nothing they can do against governmental agencies: Some consider governmental agencies (including ‘NSA’ and ‘BND’) as powerful enough that, independent of preventative actions, they would be able to track everything anyway. ‘[Using the FlashPlayer] Can be probably only used by governments with power, to abuse such data’ (4); ‘this scandal from the NSA. And I think there is much more behind it, then the spy from BND’(8). Some mentioned that it does not matter how much effort they spend ‘they will know everything about me’(19); and that in general ‘(not mentioning governmental institutions) it is difficult to protect against advanced attacks without expending enormous effort’(6) or ‘stop using the Internet’(13).

Misconceptions related to applied countermeasure: When asked which countermeasured participants applied, many mentioned measures related to careful usage of the Internet, be it ‘not opening all webpages’(1), ‘not surfing on unknown webpages’(7), or ‘automatically leave untrustworthy webpages’. Some participants believed that Firewalls protected their privacy as they ensured that sensible data would not be sent over the Internet. ‘One can, monitor what kind of data in sent from the computer. Firewalls, or so’(10). Very few believed that using Linux instead of Windows would help and some believed that not using the Internet Explorer as web browser already improved the situation. Very few proposed public computers or Internet cafes as possible solutions to avoid being tracked.

Not being aware that cookies can be deleted: Many participants were not aware that they could delete cookies. Some participants mentioned that they were aware that cookies could be deleted but had not yet done so, e.g. because they were afraid of the consequences: ‘not allowing cookies to be saved. However, the question is whether the web pages will be still functioning’(10).

Not being aware of the web browser’s privacy mode or in which scenarios it should be used: Surprisingly, half of the participants knew that the privacy mode existed. However, this also meant that some were not aware. Those who had used it did so for different reasons: surfing over unprotected Wireless LAN, watching adult content, searching for presets, when sharing the computer with others, and in order to not get tailored search results.

Not being aware of Proxies and/or Tor or what it actually does protect against: Some participants
mentioned that they had heard of proxies. However, mostly misconceptions:  ‘Changing the settings, when one cannot get to somewhere’ (2);  ‘It blocks particular pages’ (16);  ‘The connection is made through another server...the other service is just more secure that the server of the [web page] provider’ (14). Thus, in total most participants do not know when they should use a proxy service. Some mentioned that they use Proxies for reason other than privacy protection, for example to open unavailable content, or ‘to circumvent this blocking’ (14). Only some had heard the term Tor while only a few had an idea what it does (‘without leaving traces’ (8); ‘it makes a cover around the IP address’ (2)) and only the experts could explain what it actually does.  

Not being aware of more advanced privacy protecting tools: Most of the other mentioned privacy protecting tools were only known to the experts. Some participants had heard of ad blockers. Those who used it could not explain why it was a privacy protecting tool as they only used it because they were annoyed by the advertisements.  

Summary: Most effective countermeasures are not known to most of the participants and those they have heard of, it is very likely that they misunderstand the functionality. Some think they actually protect themselves but which is not the case in the context of identification and tracking on the Internet. Thus, not being aware of effective countermeasures is another explanation for not protecting their privacy against identification and tracking on the Internet.

5.6 People are not able to use them

While usability issues of different privacy protection tools have been identified in past [14], we only focused on the private mode of Firefox, Chrome, and Internet Explorer since these are the most popular browsers. Our findings are presented in this subsection.  

Most of the participants stated that it was difficult to notice whether the web browser’s private mode is active or not in particular with Firefox and Chrome. Most of the participants preferred the Internet Explorer’s implementation because the statement ‘private mode’ is displayed next to the address bar which needs to be considered anyway. Some mention that using the terms is easier to interpret than the symbols used by Firefox and Chrome. Moreover, the position of the symbol in Chrome was criticized by a couple of participants: ‘one looks only at top right, when one wants to close the window’ (1).  

Regarding the provided information after having started the private mode, many people reported about the following issues: (a) information provided by IE is misleading and not fully correct, i.e. one might think to be more private than what is actually provided; (b) statement ‘such as more Information [sowie weitere Daten]’ ‘one can read more about it, if you want to, or not?’ (10); (b) contradicting and confusing information provided by Firefox, i.e. first it is stated that one can surf anonymously but than exceptions are mentioned including that employer still know which webpages they visit; (d) impression that it is complicated to use the mode (e.g. ‘all extension deactivated. Sounds weird [in Chrome]’ (7, 10, 18)) and not clear whether full functionality is provided; (e) it is assumed that terms like cookies are known; (f) information missing in which situations usage is recommended.  

Summary: The literature, as well as our own findings for the privacy mode, show that there are a number of usability issues that keep people from using corresponding privacy protecting tools.

5.7 People side-tracked by other reasons

A couple of sub-themes were identified that do not fit to any of the other themes and are therefore listed here.  

Effort: Some participants complained about too much effort to find out which threats exist, how they can be addressed, and what the restrictions are. For instance changing to and from the private mode seems to be too complicated and error-prone.

Restricted functionality: A few participants are uncertain about whether there are functional restrictions when using the different countermeasures: ‘that pages won’t work anymore, without cookies’ (10); ‘Browser fingerprinting used’ ...probably to optimize the web pages, possibly’ (10).

Security level not clearly defined: People do not understand which threats and types of attacks are addressed by which tool; and worry that many attacks are still possible  

Tor being to slow: The experts mentioned that they tried Tor but did not to do so anymore because of the delays.

5.8 Communication Channels

Participants were more concerned after the interviews. This suggests that increasing awareness might change attitudes towards privacy protection in the context of identification and tracking. Correspondingly, it is worth knowing what participants think are appropriate communication channels. The following were mentioned (starting with those mentioned most often): Wikipedia, google search, newspapers such as (FAZ, Spiegel, Bild), forums, events (including at school and at the university), web pages from the BSI, online newspapers and large companies such as Amazon; as well as the start page of web browser. Beside pure text based information videos and games were mentioned as well as interactive learning principles. Some mentioned that a problem when searching for information is that it is necessary to know the proper terms to find the right information.

5.9 Further interesting findings

In this subsection we summarize interesting findings which do not directly related to possible explanations or communication channels.

Threats from different web service providers: We noticed that participants have different ideas of the trustworthiness of web service providers. A few think that larger ones are more likely to sell sensible data, others think that small ones are less likely. One participant even mentioned that ‘that american companies save more data...on one’ (9).

Opinions about personalized advertisements: Some participants think that personalized advertisements are useful; some are ok when they have an account while they are annoyed by personalized addvertisements provided by third parties. Some participants were of the opinion that all these adverts were useless or even annoying.

Protecting against friends and victims: A few mentioned friends to protect against (in particular if looking for presents); people using the same device; or are sitting next to each other; or ex-friends.

Demanding more transparency: Some participants talked about transparency, e.g. demanding more transparency from providers in terms of stating clearly what data is collected, how, and the purposes it is used for, as currently
data is collected without informing users properly; people also mention that they get the impression that services and tools do not really improve transparency.

5.10 Limitations

The limitations of our research mainly lie in the interview group: Besides the fact that mainly younger participants joint as well as many students, most of our participants were recruited by the computer science friends and colleagues. Hence, the identified explanations is mainly representative for young people (mostly academics) while for the completely picture a broader audience needs to be interviewed.

6. CONCLUSION AND FUTURE WORK

Our goal was to elaborate possible explanations for why people do not act to prevent electronic surveillance and thereby preserve their privacy. We identified seven different types of explanations that need to be addressed in order to change the situation, i.e. encourage the use of effective countermeasures. This shows that the myth of unusability impacting the deployment of counter measures is by far not the only reason for the slow uptake of privacy protecting tools. The interviews revealed a couple of misconceptions in particular with respect to the difference between explicit and implicit sensible data; as well as between privacy and security as a concept. The findings also indicate that security, in terms of computers not working (or working to slow) or losing money, is more concrete and easily brought to mind than privacy related threats. Furthermore, most participants could only think of personalized advertisements as consequence of being identified and tracked.

In general, the evaluation of the interviews indicates that there is a big gap between passive and active knowledge about privacy-related threats and which countermeasures mitigate against each threat. Furthermore, people seem to be less motivated to protect against privacy-related threats because of the lack of awareness of the potentially negative consequences.

In order to improve the situation, we recommend improving the communication related to threats, corresponding consequences, and effective countermeasures (including their restrictions) using popular web pages, newspapers and forums. It goes without saying that the usability of the different privacy protecting tools still needs to be improved but doing so will not necessary, on its own, lead to a greater uptake of these tools.

Acknowledgements

We would like to thank everyone who supported us by participating in the user study. We also want to thank Karen Renaud for her useful comments and we thank CASED and EC SPRIDE for supporting this work.

7. REFERENCES


