

Wirtschaftsuniversität Wien

Magisterarbeit

Titel der Magisterarbeit:
Web Browsing Privacy Threats and Standards

Verfasser: Andreas Badelt, BSc (WU)
Matrikel-Nr.: 0450459
Studienrichtung: WINF-M03
Textsprache: Englisch
Beurteiler: Univ.Prof. Dipl.-Ing. Mag. Dr. Wolfgang Panny
Betreuer: Dipl.-Ing. Mag. Dr. Albert Weichselbraun

Ich versichere:

dass ich die Magisterarbeit selbständig verfasst, andere als die angegebenen Quellen und Hilfsmittel nicht benutzt und mich auch sonst keiner unerlaubten Hilfe bedient habe,

dass ich dieses Magisterthema bisher weder im In- noch im Ausland (einer Beurteilerin/ einem Beurteiler zur Begutachtung) in irgendeiner Form als Prüfungsarbeit vorgelegt habe,

dass diese Arbeit mit der vom Begutachter beurteilten Arbeit übereinstimmt.

Datum

Unterschrift

Web Browsing Privacy Threats and Standards

Andreas Badelt, BSc (WU)

While surfing the Internet, users often provide data to customize Internet-applications or to surf more comfortably. They also leave data behind without their explicit knowledge just by browsing a website. This data can be processed to gain more information about users once it has been provided, unwillingly or not.

Obviously, there is often a certain trade-off between functionality and privacy on the Internet and although data protection laws are in place, these laws do not necessarily protect data in a way users would expect. Now, considering the amount of user-data which is collected on the one hand and the lack of understanding how data can be used once it has been provided on the other, the risk of data-misuse arises. That is why privacy standards and privacy tools are important to dispel privacy threats or at least to reduce privacy risks when browsing the Internet. However, as of yet, no such standard is available which is both widely accepted by website operators and supported by the software industry. This thesis not only addresses the issues mentioned above, it also provides an overview about the requirements for privacy tools, introduces popular privacy applications and describes the implementation of a privacy plugin for the Firefox Internet browser.

Contents

Abbreviations	vii
Listings	viii
List of Figures	ix
List of Tables	x
I. Theory	1
1. Introduction	2
1.1. Objective	2
1.2. Research question	3
1.3. Thesis structure	4
2. Privacy Threats	6
2.1. Individual privacy attitudes	6
2.2. General privacy on the Internet	7
2.2.1. Financial privacy	7
2.2.2. Medical privacy	8
2.2.3. Political privacy	8
2.3. Privacy sensitive applications	9
2.3.1. Personalization	9
2.3.2. E-Commerce	10
2.3.3. Search engines	11
2.3.4. Social networks	12
2.4. Privacy sensitive technologies	14
2.4.1. Cookies	14
2.4.2. ISP: tracking of data traffic	15
2.4.3. Online profiling	16
3. Privacy Standards	18
3.1. The Platform for Privacy Preference Project (P3P)	18
3.1.1. P3P - an introduction	18
3.1.2. Requirements	19
3.1.3. P3P policy reference files	22

3.1.4.	P3P policies	24
3.1.5.	A P3P Preference Exchange Language (APPEL)	31
3.1.6.	Existing P3P user agents and software tools	31
3.1.7.	Future of P3P	32
3.2.	The Enterprise Privacy Authorization Language (EPAL)	32
3.2.1.	EPAL - an introduction	33
3.2.2.	Requirements	33
3.2.3.	EPAL policies	33
3.2.4.	EPAL vocabularies	34
3.3.	The eXtensible Access Control Markup Language (XACML)	35
3.3.1.	XACML - an introduction	36
3.3.2.	Requirements and structure of XACML	36
3.3.3.	Summary	38
4.	Requirements for Privacy Tools	39
4.1.	Generic software quality requirements	39
4.1.1.	Functionality	39
4.1.2.	Reliability	40
4.1.3.	Usability	40
4.1.4.	Efficiency	40
4.1.5.	Maintainability	40
4.1.6.	Portability	41
4.1.7.	Quality in use	41
4.2.	Privacy specific requirements	41
4.2.1.	Support of privacy standards	41
4.2.2.	Prevention of threats	41
5.	Evaluation of Existing Privacy Tools	43
5.1.	Browsers	43
5.1.1.	Microsoft Internet Explorer	43
5.1.2.	Firefox	46
5.1.3.	Safari	46
5.1.4.	Opera	46
5.2.	Plug-ins	47
5.2.1.	Adblock Plus	47
5.2.2.	NoScript	47
5.2.3.	Flashblock	48
5.2.4.	BugMeNot	48
5.2.5.	TrackMeNot	49
5.2.6.	Private Web Search (PWS)	49
5.3.	Proxies	50
5.3.1.	Anonymizer	50
5.3.2.	The Onion Router (TOR)	50
5.3.3.	Switchproxy	51

II. Applied part	53
6. Development of a Privacy Plug-In	54
6.1. Add-on development for Firefox	54
6.1.1. The extension concept	54
6.1.2. Developing Firefox extensions	55
6.1.3. File structure	55
6.1.4. Flexible graphical user-interfaces (GUIs)	56
6.1.5. XUL templates	56
6.1.6. Accessing SQLite databases	57
6.1.7. Accessing the local filesystem	57
6.1.8. Using an integrated development environment (IDE)	57
6.1.9. Debugging Firefox	57
6.2. Initial situation	58
6.2.1. Compatibility	58
6.2.2. Functionality	59
6.3. Up-port and enhancement of a P3P add-on	60
6.3.1. Compatibility	60
6.3.2. Functionality	62
6.3.3. Issues	65
6.3.4. Evaluation of the developed add-on	66
7. Conclusion	69
References	71
A. microsoft.com's P3P policy	79
B. E-Mail correspondence	84

Abbreviations

APPEL	A P3P Preference Exchange Language
EPAL	Enterprise Privacy Authorization Language
FF	Mozilla Firefox
HTTP	Hyper Text Transfer Protocol
IDE	Integrated Development Environment
IDN	Internationalized Domain Name
IEEE	The Institute of Electrical and Electronics Engineers
IM	Instant Messaging
ISO	The International Organization for Standardization
ISP	Internet Service Provider
MDC	Mozilla Developer Center
MSDN	Microsoft Developer Network
MSIE	Microsoft Internet Explorer
P3P	The Platform for Privacy Preference Project
TOR	The Onion Router
VPN	Virtual Private Network
W3C	The World Wide Web Consortium
WWW	World Wide Web
XACML	Extensible Access Control Markup Language
XML	Extensible Markup Language
XUL	XML User Interface Language

Listings

3.1. HTTP response header with P3P (Source: [W3Ca])	20
3.2. Link tag method for indicating policy reference file (Source: [W3Ca]) . .	20
3.3. Policy reference file (Source: [W3Ca])	22
3.4. Policy reference file using wildcards	23
3.5. Policy reference file using the HINT element (Source: [W3Ca])	23
3.6. Policy reference file using the COOKIE-INCLUDE and COOKIE-EXCLUDE elements (Source: [W3Ca])	24
3.7. microsoft.com's P3P policy - general assertions	25
3.8. Policy statement regarding a e-mail comment form (Source: [Cra02]) . .	30
3.9. Basic XACML request example (Source: [OAS05])	37
A.1. microsoft.com's P3P policy (Source: microsoft.com)	79
B.1. E-mail correspondence with amazon.com's customer service	84
B.2. E-mail correspondence with Microsofts's customer service	85

List of Figures

5.1. MSIE: privacy settings	44
5.2. MSIE: privacy report	45
5.3. MSIE: Privacy Bird	45
5.4. Safari: security settings	46
5.5. Opera: cookie settings	47
5.6. Flashblock	48
5.7. BugMeNot	49
5.8. Java Anon Proxy (JAP)	50
5.9. The Onion Router (TOR)	51
5.10. Switchproxy	52
6.1. Privacyfox: access translated P3P policy	59
6.2. Privacyfox: display translated P3P policy	60
6.3. Webprivacy: options	63
6.4. Webprivacy: P3P policy does not match	64
6.5. Webprivacy: P3P policy does match	65
6.6. Webprivacy: installation	67

List of Tables

3.1. PURPOSE sub-element definitions and translations (Source: [W3Ca]) . . .	28
3.2. RECIPIENT sub-element definitions and translations (Source: [W3Ca]) .	29
3.3. RETENTION sub-element definitions and translations (Source: [W3Ca])	29
3.4. An example EPAL privacy rule (Source: [IBM03])	34
3.5. An example EPAL request (Source: [IBM03])	34
6.1. Five states of Webprivacy	63

Part I.
Theory

1. Introduction

This thesis addresses the users' need for privacy on the World Wide Web (WWW), common threats to privacy and which solutions exist to encounter these threats on the Internet¹.

Many Internet-applications such as e-commerce shops, online banking applications or search-engines need certain input to function and this data is often provided by the users of a certain application. This collected data is protected by laws and regulations. However, these laws do not necessarily protect the data as strictly as users would expect. If users have a lack of understanding of how their data can be (mis-)used once it has been collected, then this leads to a precarious situation considering the false prospects of data protection laws.

1.1. Objective

The issue of web browsing privacy and privacy threats is especially important as more and more services are delivered online: Internet banking, medical advice, e-mail, tax return, e-learning, shopping and much more [Ant07]. All these services need certain user information to operate. A web shop for example needs an e-mail address to send status e-mails to the customer. It needs credit card information or bank account information to withdraw the money. And it has to know the real name and address of a user to deliver the products purchased. For the convenience of the customer, all this data can be stored on one of the vendors' servers - if the customer has provided this information once, he or she does not have to provide it during the checkout for the following purchases. From a customer perspective, that definitely is convenient.

On the other hand, does the user still have control over his data? Can he or she be absolutely certain, that no one is going to mis-use the data provided?

Certainly, there are regulations and laws which, in most countries, prohibit acts such as mis-use of personal data (e.g. "The Federal Act Concerning the Protection of Personal Data" - DSG 2000 - in Austria or the "Data protection directive" 95/46/EC of the European Union). At least the owner of the data has to agree if his or her data is being used for other purposes. Privacy policies, for example, can be found at every commercial website or portal. However, the question is how many web users are actually able to find such policies. And if they have been found, if they are read and also understood.

Another issue arises when it comes to the lack of understanding how information and data provided during web browsing can easily be found on the web. If one takes a closer

¹Although technically not correct, the terms "WWW" and "Internet" are going to be used interchangeably in this thesis for the readers' benefit

look at the behaviour of users of social networks such as Facebook or MySpace, it can be questioned if all users are aware of the fact that their data can be found over search engines such as Google, Yahoo! or MSN. Of course, most platforms provide privacy settings to hinder web-crawlers to index such information and prohibit other users of such platforms to access it without prior consent of the owner. However, studies show that most users do not know about such issues and platform providers do not necessarily have the most stringent privacy settings in place by default [Jon05]. A lot of web users “[...] are posting content online without thinking about the electronic footprint they leave behind” [Gua07] which can have severe consequences in the future - think about employers “googleing” you before inviting you to a job interview or parents taking a look at the latest party pictures of their children.

Methods such as “profiling” and “tracking” enable search engines, web shops and other companies to create user profiles [Awa06]. With such user profiles, all kind of preferences can be found out and certain conclusions can be drawn. If a user at amazon.com for example is mainly searching for English science fiction books on Saturday nights, amazon.com (or their algorithms, respectively) could come to the following conclusions:

- The best way to make sure the advertising on the website has a high conversion rate is to show advertisements to the user which are related to science fiction
- There will be a higher chance of selling the user another book if it is similar to the books which were sought for
- The user may not have a lot of social life because he is regularly searching for books on a Saturday evening

Although the last conclusion may seem a little bit far out, search engines are able to collect and connect millions of user profiles. Privacy scandals such as AOL’s publication of search queries plainly show us how much personal information such companies can generate by connecting information: although the data of AOL was anonymized, everyone with access to the Internet was able to find out that “AOL user 311045 apparently owns a Scion XB automobile in need of new brake pads that is in the process of being upgraded with performance oil filters” or that AOL user 005315 “searched for information about prison inmates, gang members, sociopaths in relationships, and women who were murdered in southern California last year” [CNT06]. In a very dramatical way, such examples highlight the need for privacy on the Internet and the need for ways to make privacy policies easy to understand for (unexperienced) web users.

1.2. Research question

The thesis addresses the question of how Internet users can make sure that their data is only used with their knowledge for the purposes they know and approve of. To answer this question, threats to privacy when browsing the web have to be discussed as well as common web privacy standards.

The topic of this thesis also specifies a certain need for an applied point of view. For that reason, requirements for web privacy tools are going to be introduced and such tools will be evaluated considering these requirements.

In addition, the applied part of this thesis will focus on up-porting² and enhancing an existing add-on³ for Mozilla Firefox which adds a Platform for Privacy Preference Project (P3P) functionality to the open source web browser which feature it is currently lacking.

Taking everything into account one can say that this master thesis has two goals:

1. Based on literature and examples, the topic of privacy while browsing the Internet should be addressed, threats to privacy should be highlighted and privacy tools should be evaluated .
2. As a starting point for the Mozilla community to support P3P, an add-on for Mozilla Firefox should be developed that anticipates the current lack of P3P support of Firefox.

1.3. Thesis structure

The content of this thesis are web browsing privacy threads and standards, that is which threats to privacy exists while surfing the Internet and which standards were developed to hinder such threats. Already existing tools are going to be evaluated and an existing tool is going to be enhanced as the applied part of this thesis.

Chapter two, **2 Privacy Threads**, deals with privacy threats on the Internet. In particular, general privacy issues are going to be discussed as well as common applications which require personal data such as e-commerce applications or social network platforms. In addition, some (technical) background information about privacy on the Internet will be provided, especially about cookies, Internet Service Providers (ISPs) and methods for user tracking and profiling.

The next chapter, **3 Privacy Standards**, introduces three extensive privacy standards which were developed by organizations or norming institutions: W3C's P3P, IBM's Enterprise Privacy Authorization Language (EPAL) and OASIS' eXtensible Access Control Markup Language (XACML). As parts of this thesis deals with the implementation of P3P, the emphasis in this chapter lies on the P3P Standard.

In chapter four, **4 Requirements for Privacy Tools**, generic software quality requirements (based on ISO/IEEE standards) are going to be discussed as well as privacy specific requirements according to literature.

²“Up-porting” describes an activity in software engineering where a certain piece of source code is adopted to a changed (software) environment

³For the readers' benefit, the term “add-on” and “extension” will be used exchangeable in this thesis although strictly seen “extension” is correct

The fifth chapter, **5 Evaluation of Existing Privacy Tools**, deals with the actual evaluation of privacy tools based on the requirements defined in chapter 4. As there are several kinds of privacy tools, this chapter will focus on tools used in an Internet environment such as browsers, browser add-ons and proxies.

In the last chapter, **6 Development of a Privacy Plug-In**, the development of a privacy add-on for Mozilla Firefox is described. This chapter is, in combination with Appendix A and B, the documentation of the applied part of this thesis. Besides a short introduction to add-on development for Mozilla Firefox, an overview about the existing add-on and the enhanced extension will be provided. Finally, the new add-on is going to be evaluated considering the defined requirements.

2. Privacy Threats

Privacy is the “claim of individuals to determine for themselves, when, how and to what extent information about them is communicated to others” (Westin 1967 cited in [Oli04]). When it comes to privacy on the Internet, there are a lot of threats which menace this claim and they can be categorized in different ways. According to [Sar03] there are general threats which menace financial, medical or political privacy. Then there are common privacy sensitive applications which are used by Internet users quite frequently such as search engines, social network platforms, e-commerce applications or personalized web-services. All these applications have one thing in common: they use sensitive user-data to deliver their services. That is why common privacy threats and privacy sensitive applications are going to be introduced in this chapter. Additionally, some background information will also be provided on how technical issues play an important role in privacy on the Internet.

2.1. Individual privacy attitudes

When discussing privacy threats on the Internet, the biggest threat is sometimes forgotten: the person in front of the computer. All kinds of technical solutions can be implemented and enforced but will not make a difference if users in front of the computer do not reflect about their actions. This topic was already an issue when the World Wide Web was becoming more popular. In 1998, Alan Westin published a categorization of users and their individual attitudes to privacy as also described in [Kob07]. These three different attitudes can also be applied to the Internet and are described as the “privacy fundamentalists”, the “privacy pragmatists” and the “privacy unconcerned”:

- **Privacy fundamentalists** can be characterized by their “extreme concern about any use of their data and an unwillingness to disclose information, even when privacy protection mechanisms would be in place” [Kob07].
- **Privacy pragmatists** on the other hand are generally concerned about their privacy, but not as much as the fundamentalists. If they can see a certain reason for disclosing private data, they are willing to do so.
- The exact opposite of the privacy fundamentalists are **privacy unconcerned** which are not as distressed by the thought of disclose their private data and how it may be used by third parties.

Despite individual privacy attitudes there are other circumstances where privacy may be threatened. That is why in the following two sections, general privacy issues on the

Internet and areas of common Internet applications which use (sensitive) personal data will be introduced.

2.2. General privacy on the Internet

This section will deal with general privacy issues on the Internet. For that reason a common approach to categorize personal data as described in [Sar03] will be used. The three main categories are financial, medical and political privacy. Although all categories are generic, their description in this thesis is restricted to the World Wide Web and should introduce these issues based on examples. Therefore, financial privacy will be discussed on the examples of fraud and identity theft, medical privacy by discussing the transfer of patient records over the Internet and seeking medical help online and political privacy will be discussed by highlighting privacy issues in e-government and free speech on the Internet.

2.2.1. Financial privacy

When it comes to financial privacy, two threats are very common on the Internet: fraud (such as in online banking and phishing¹) and identity theft. According to [Sin07] a huge rise in online banking fraud could be identified in the last years which mainly is due to two reasons: First of all, Internet access has become very cheap in most industrialized countries and banks try to get their customers to use online banking for obvious reasons: For clients, it is a very convenient way to conduct banking business. For banks, it is a way to cut costs on a double-digit scale [Tan00]. That means that there are a lot of online banking accounts which can be attacked. The second reason is the growing number of phishing mails because there are more and more users with e-mail addresses available. In most cases, these e-mails look like official e-mails, for example from the customers' bank, asking them to update their contact details. By including a link which leads to a fake website (very often a copy of the original website) it is made sure that users send their sensitive data to the "right" party. In the best case this data (such as address-data) is sold, more often account data is requested and used to debit money from a banking account. However, nowadays banks are aware of this problem and try to inform their customers as much as possible about such phishing mails. Additionally, new security features are being introduced to make phishing harder and therewith protect customer data [Hil06].

Identity theft is closely connected to online fraud whereas [Kah08] defines it as "the malicious use of personal identifying data" and provides several types of identity theft: new account fraud, existing account fraud and friendly fraud. According to [Kah08] **new account fraud** is characterized by the use of personal data such as the social insurance number, date of birth and address to open a new account. By **existing account fraud**

¹"Phishing" describes a methodology to acquire sensitive information such as credit card numbers by sending users fake e-mails with links to fake websites which try to bring users to enter their sensitive data

the theft of credit cards or other transactional account data is described whereas the term **friendly fraud** relates to legitimate purchases where - after the transaction has been concluded - the purchaser denies the legitimacy of the debt. However, the field of study on identity theft is huge and other categorizations have been provided as well [New05].

2.2.2. Medical privacy

Other issues arise when it comes to threats to medical privacy. Nowadays it is common to surf the Internet to find help or medical advice via forums or other websites. Research shows that there is concern about privacy when it comes to innovative concepts such as telehealth where new technology such as e-mails, web cams or forums are used to communicate [Dro06].

Another question addresses the topic of sharing patient records over the Internet and how to make sure that confidentiality is guaranteed in such cases. [Rin97] describes the still up-to-date topic of how to make sure that transferring patient records over the World Wide Web does not have a negative impact on confidentiality. Although today all different kinds of encryption technologies are available, the problem lies not only in the actual transfer. It is rather a problem of access at the starting and destination points of the transfer and how to make sure that just authorized personnel can access electronic patient records [Rin97].

2.2.3. Political privacy

Today, all kinds of services in the private sector are delivered online. However, there is also a growing pressure to deliver public services via the Internet [Sid08]. The government of Estonia for example has one of the most sophisticated approaches to e-government in place [Kit08]. This is the outcome of a development which started in 1996 where Internet banking was successfully introduced in Estonia. As government agencies implemented the same authorization system known from online banking in their applications, it pushed the introduction of other e-government services. As a consequence, Estonians, for example, are able to file their taxes online since 2000 [Kit08]. However, Estonia is of course not the only country which facilitates e-government. Another important issue in e-government is e-voting such as for example discussed in the United States of America. Among other things, an e-voting system for absent military and oversea citizens has yet not been implemented due to (among other things) arising privacy and security concerns [GOA07]. Although e-voting is probably the most sensitive application of e-government, all services need sensitive data of its citizens to operate properly - and are therefore a potential privacy risk due to misuse of the data provided by third-parties or data collection by governments. That is why privacy issues have to be addressed seriously in this context [Bel06].

The topic of political privacy should also be addressed in a more broader sense. Free speech for example is a foundation stone of every democracy [Cel08]. However, it is,

for example, not unproblematic to be a government critical blogger² in all countries. Taking a look at the censorship in the People’s Republic of China or the limited access to the Internet in China the question has to be raised how opinions can be published without the fear of imprisonment or other sanctions [Pau07]. Privacy is the ultimate goal here. In such an environment, privacy can only be achieved by using technology (such as anonymous proxies or virtual private networks (VPN)) to stay anonymous while browsing the web and publishing content on the web [Opp05].

2.3. Privacy sensitive applications

After introducing threats to financial, medical and political privacy, common privacy sensitive applications should be introduced. Such applications use personal data to perform a certain service to the user. Therefore, popular services on the Internet that use such sensitive data will be discussed and arising threats will be highlighted. The following categorization of web-applications is based on [Kob07].

2.3.1. Personalization

Personalization has a different meaning in different industries. As this thesis is focused on the World Wide Web, the definition of the Personalization Consortium (cited in [Ves07]) will be used:

Personalization is the use of technology and customer information to tailor electronic commerce interactions between a business and each individual customer. Using information either previously obtained or provided in real-time about the customer, the exchange between the parties is altered to fit that customer’s stated needs as well as needs perceived by the business based on the available customer information.

This definition was chosen for two reasons: first of all, it explains what personalization is. But more importantly, it emphasizes the need of data to actually provide personalization. Newspaper websites for example can often be adapted to one’s needs - if wished, they will only display sports articles, the daily cartoon and articles on foreign affairs. Of course, this is very useful but on the other hand this information can be used to create profiles about users. If a website “knows” what articles users read, it may conclude which preferences users may have. Certainly, equipped with this knowledge, it will only display advertisements which fit to users’ preferences [Miy08]. However, most big websites offer some personalization to its users [Kob07]. Taking a look at one of the most popular search engines - Google. With “iGoogle” users can customize the start page of Google to their specific needs. Users can not only select their preferred language, they can display selected news, weather forecasts and other gadgets such as Sudokus. Taking an isolated look at every service does not reveal any threat to privacy.

²A “blogger” is a person who regularly posts (personal) information on a website in journal style (a “blog”)

But if users access their news over iGoogle, search Wikipedia via iGoogle and check the weather forecasts via iGoogle, Google may get to know this users better in terms of their preferences. But this is just the beginning. If users check their Google Mail account via iGoogle, post memos and “to do” lists on iGoogle, send text messages via the free text messaging service and calculate the route to their next business partner via iGoogle, then this may lead to a huge amount of data which could be collected about these users (and hence hidden knowledge could be extracted from these multiple sources with a technique called “data mining” - see Section 2.3.3). Although a lot of these services are delivered by third parties (such as Map24.de) it is arguable if the normal Internet use can distinguish between embedded services on a portal and services delivered by Google. Such an example shows that there is a certain trade-off between privacy and personalization. As described in [Awa06] “consumers who value information transparency features are less willing to be profiled online for personalized service and advertising” which leads into two directions: either (I) get users to provide personal data by informing them about the benefit of personalized features, or (II) develop technologies which allow companies to provide personalized services on the Internet without having to collect too much personal data such as a privacy enhanced search-engine. However, it may be questioned if such technologies will be implemented on a large scale because personal data of customers is a valuable asset to companies [Xu07].

2.3.2. E-Commerce

Personalization and customization are also closely linked to e-commerce. One classical example of e-commerce is amazon.com. Registered users at amazon.com (or one of the equivalent country websites) will have noticed that when displaying a product, amazon always suggests similar products. This is because of amazon.com’s “Customers Who Viewed This Item Also Viewed” feature. In addition to that, amazon.com recommends similar products based on user’s previously purchased articles [Gre03]. Such features certainly are nice and useful to customers but they are also a threat to privacy as users get profiled. Users are partly aware of these issues and research shows that they share concerns about the disclosure of personal data in e-commerce [Nam06]. As a consequence, gaining the trust of users is a vital topic in this field. This is especially true when it comes to data which may be used in marketing activities [Moo05]. Studies show that confidence and trust in a website do not only play an important role, they can also be influenced by the publishers of the website: by providing well-known logos, reliable external trust certificates and by a convenient and easy to use website, organizations can make users feel more secure and hence provide sensitive data more easily [Nam06]. However, such methodologies must be questioned, especially when it comes to trust certificates. There, recent studies show that users may be more guided by the actual design of the trust certificate logo than the organization which awards such certificates and what such certificates stand for. This leads to the situation were fake trust certificates were recognized more often than real ones [Moo05].

However, privacy can also be “a seller”. [Gid06] researched on the question if “availability of comparison information about the privacy practices of online merchants affects

users behavior” and found that “[...] when privacy policy comparison information is readily available, individuals may be willing to seek out more privacy friendly web sites and perhaps even pay a premium for privacy depending on the nature of the items to be purchased” [Gid06].

2.3.3. Search engines

When it comes to search engines, privacy plays an important role because for Internet users they are a vital part of their daily routine. In this context the threat to privacy shows different characteristics:

- Personalization and customization as already described in the sections above
- Creation of user profiles for improving search quality
- General data mining capabilities of search engines

The question of personalization and customization has already been discussed above, based on the example of iGoogle. However, there are more threats to privacy when it comes to the usage of search engines. One such threat is the creation of user profiles for improving search quality. [Xu07] describes an experiment where it was shown that providing personal, sensitive data improves the quality of search results. This is due to the fact that by providing information about one’s preferences or other characteristics the search engine is able to filter unsuitable search results. If users are male and under 30 years of age for example and they provide this information, they will probably get better search results if searching for swimsuits simply because the search engine can filter out all websites which provide information about female swimsuits. Once provided, this information is probably stored for an undetermined amount of time. However, users are able to deny such information if they want to.

Despite creating user-profiles, there is also another technology which, if mis-used, can pose a threat to privacy called data mining. [Han01] defines data mining as “the analysis of [...] observational data sets to find unsuspected relationships and to summarize the data in novel ways that are both understandable and useful to the data owner”. In the case of search engines this leads to incidents such as the scandal where AOL published a database of anonymised search records for scientific use. Unfortunately, they were not used for scientific purposes only which, shortly after the publication, led to whole user profiles being published on the Internet [CNT06]. Although such big incidents rarely occur, they clearly highlight the capabilities of search engines to collect the data of its users and, once processed and interpreted, how such data can be abused. However, these threats are not only applicable for the actual users of search engines, but also for everyone who published content on the Internet without proper knowledge of its technologies such as web-crawlers³. These web-crawlers collect data by indexing billions of websites if not explicitly locked out from the website. Over time, search engines can

³Web-crawlers are programs sent by search engines to “crawl” from website to website to index them and hence make them available in the database of a search engine

collect a lot of personal data about people by indexing several websites and hence make this sensitive data available to anyone with one click by entering a simple query into the search engine - security through obscurity is not existent anymore if personal data has been published on the Internet [Pik2006]. This problem area especially arises when it comes to social networks which the following section is going to deal with.

2.3.4. Social networks

Social networks such as Facebook, LinkedIn, MySpace or StudiVZ are getting more popular every day, hence attract more users and collect more data [Gro05]. One principle of such platforms is that users have to register to create a profile. There they can enter all kind of personal data: Name, address, contact information such as e-mail or telephone number, date of birth, hobbies, photos, relationship status, employer and so on. If users want to, they are literally able to publish their whole curriculum vitae. One of the main reasons to join such a social network is to get connected to friends and other people by “adding” them as “friends”. With this feature, social networks can be calculated by the provider of the platform: friends of friends are suggested to other users because they may know them, live near them or attended the same class, different connections to other friends can be displayed and so on. All this leads to a huge threat to privacy if not necessary mechanisms are in place to make sure that all this very sensitive user data is not mis-used (e.g. for data mining), sold (e.g. to direct marketing companies) or stolen (and then, for example, used for identity theft). Although all such platforms offer privacy settings, the default settings are not very strict. The reason for this was already mentioned earlier: there is a trade-off between privacy and functionality, especially for social networks. In addition to that, every information about its users is an asset for the platform operators. [Gro05] discusses privacy implications of social networks based on the example of Facebook which are broken down into the following categories:

- **Stalking:** As user profiles often not only include information about the residence but also about the occupation (such as the university and currently attended classes at university), stalkers can easily identify the physical location during the day. Especially status messages can also reveal information which could be used to locate the physical location of users. Besides physical stalking, potential adversaries could stalk their victim via instant messaging (IM) as IM account-names are very often provided in user profiles.
- **Re-identification:** Besides using data such as date of birth and zip-codes, users can easily be identified via posted pictures. Depending on the country, available data on social networks can also easily be used for identity theft. The social security number in the US, for example, can almost be estimated if the following information is available: date of birth, hometown and current residence - all this information is often present in user profiles [Gro05].
- **Building a digital dossier:** By collecting data about users, a digital dossier can easily be build, especially over time. Changing home-towns, partnership details,

jobs or friends can be monitored and compiled to a report. One not necessarily has to actively monitor accounts to receive information about the users' past - this, at least in the case of Facebook, is already done for other users by the platform itself. Another source of information are users' pictures. Taking a look at the pictures available often provides enough information if users have been registered for a certain amount of time or pictures about users have been uploaded. Potential employers could, for example, check out applicants by taking a look at their pictures, reading posts of friends and so on.

- **Spear phishing:** [Jag07] highlights the issue of spear phishing or context aware phishing. In such phishing attacks, personal information is used to adapt the attack specifically to the phishing victim for example by using data available on social networks. The study shows that when exploiting social network data, the success rate of phishing attacks is four times higher than normal [Jag07]: social networks such as Facebook or MySpace can easily be automatically crawled and reliable data about social networks can be extracted and stored in a database. This information can then be used to “personalize” phishing attacks for example by spoofing e-mails to make them appear to be of a friend. As the phishing victims recognize the (fake) sender (it looks like the e-mail has been sent of one of his friends, e.g. from Facebook), the user is more likely to click on the link and provide sensitive information [Jag07]. The study also supports the statement that there is a lack of understanding that data posted on social network sites is public information - and how easily this information can be abused.

Just recently, Facebook tried to implement “changes to its contract with [its] users that had appeared to give it perpetual ownership of their contributions to the service” [NYT], including all personal data and status messages entered. After a huge outcry of the Facebook community, these changes were reverted. However, it is a clear example how data collected by such social network is potentially dangerous.

However, most of these very real threats to privacy are only made possible because users are not aware about the above discussed consequences if they do not protect their data by using existing privacy settings [Jon05]. This leads to an alarming discrepancy between the perception of users (“nobody will ever find this information or is interested in it”) and the reality: not do only other people actively search for information, also automatic data collection about specific persons is taking place. By crawling the Internet, search engines provide a basis for services such as 123people.at which displays publicly existing data of specific users on the Internet. However, this issue is not only restricted to social networks. Also other popular services like blogs can pose a huge threat to privacy [Kha06].

All the above mentioned threats and application use certain technologies and methodologies to fulfill their purpose. Especially the issues of cookies, Internet Service Providers, profiling and logging is of interest and will be discussed in the next section.

2.4. Privacy sensitive technologies

The following section provides (technical) background information about the most important privacy relevant technologies and methodologies used on the Internet.

2.4.1. Cookies

A cookie is a small piece of information saved on the hard-disk of a computer. For security reasons, a cookie can only be read by the website that set the cookie on the users' computer, that is cookies which were set by "google.com" can only be read by "google.com". It can be distinguished between first- and third-party cookies. A first-party cookie is a cookie which is set from a website which the users is currently visiting, whereas a third-party cookie is a cookie which is set by a third-party (such as an advertisement-provider) on a domain by displaying an external image or the like (e.g. when browsing "nyt.com" a cookie is set by "doubleclick.com" by displaying an advertisement on "nyt.com") [W3Cd].

2.4.1.1. Purpose

Cookies were developed due to technical restrictions of the Internet protocol HTTP⁴ which is stateless, that means that after a request has been successfully processed, there is no connection between the client and the server anymore [W3Cd]. This can lead to problems with some Internet applications such as shopping carts: to successfully shop on a website, the website has to know which users put which products into their shopping carts. However, this is not possible without knowing the session-IDs⁵ of the users. That is where cookies come into play. Cookies are (among other things) used to store session information and user preferences on a computer [Coo].

2.4.1.2. Collected data

Although cookies cannot be used to steal information from a computer, they can collect data which was provided during a session on a website. If users, for example, provide user names and e-mail addresses on a form on a website, these data can be stored in a cookie and entered into the form if users visit this website again [W3Cd].

2.4.1.3. Threats to privacy

This usage of cookies definitely makes sense, for example to store users' preferences on a website. On the other hand, it can also be used for other purposes which pose a threat to privacy: tracking users. Although a cookie must always be restricted to a domain-name (such as google.com) and therefore cookies cannot be read by other websites, third-party

⁴HTTP stands for Hyper Text Transfer Protocol and takes care of requesting and sending information from clients and servers

⁵A session-ID is a unique combination of characters to identify users on a website

cookies can be created. Advertisement networks such as DoubleClick provide advertisement banners on websites. To make sure that the best possible conversion rate⁶ is achieved, such advertisement networks want to make sure that the advertisements displayed are customized to the users' interests and preferences. Applying cookies to simply track which websites were visited by users facilitates the customizing of advertisements [Miy08]. In an extreme case, the users' browsing behaviour could not only be tracked but also a profile could be created.

So although cookies can impose a threat to privacy, they are an integral part of the World Wide Web as a lot of web-applications such as net banking, online-shopping and other websites can only properly function with cookies enabled on the clients' side [Hor05]. On the other hand, browsing the Internet anonymously is not possible if using cookies [Woo06] - again, we find a trade-off between privacy and functionality on the Internet.

2.4.2. ISP: tracking of data traffic

As the name suggests, Internet Service Provider (ISP) provide the infrastructure of the Internet. They lay cables not only directly to the computer of the end-user but also physically into the earth to connect their network (and herewith their customers) to the Internet.

2.4.2.1. Purpose

Before a users' request reaches its destination, it has to go through multiple ISP infrastructure by some means or other, mostly network devices such as routers. As a consequence, ISP's can track and collect **every** single piece of data sent from or to their customers and are also able to analyze traffic-flows [Sen04].

2.4.2.2. Collected data

Especially since 9/11, government interference with storing user data has been rising [Dow05]. In the European Union for example, Directive 2006/24/EC deals with "[...] the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks [...]" [EU06]. According to this directive the following data has to be retained by Internet Service Providers for 6 months to 2 years depending on the data:

- Data necessary to trace and identify the source of a communication
- Data necessary to identify the destination of a communication
- Data necessary to identify the date, time and duration of a communication

⁶Basically, the conversion rate is a measure of how many people fulfilled a defined goal (such as buying something) after clicking the advertisement

- Data necessary to identify the type of communication
- Data necessary to identify users communication equipment or what purports to be their equipment
- Data necessary to identify the location of mobile communication equipment

2.4.2.3. Threats to privacy

Exactly this ability to track and trace every action of users on the Internet poses a threat to privacy as users are not able to make use of the Internet in an anonymous to access information way as often proclaimed [Woo06]. This is especially true as ISP nowadays have to store even more information about their customers as described above [Dow05]. This collected data can technically be used in all kind of ways. For example, when users visit websites, these visits are normally logged by the webserver which processes the requests necessary to display the website in the customers' browser. Every single access is logged in the logfiles of the webserver including the IP-address⁷ of the user [NLC]. Now the IP-address is the piece of information which links the data collected in the logfiles of the webserver with the data collected of the customers' ISP. So technically, every user can be personally identified if connecting these two databases, provided that this data is logged. With these data retention policies in place, specific users can be kept under "online surveillance" by government agencies [Dow05].

These issues show the significance of this topic and that (online) privacy will become an even more important topic in the future.

2.4.3. Online profiling

Today, "online profiling is the collection of information about Internet surfing behaviour across many different websites for the purpose of formulating a profile of users habits and interests" [Wie02], especially in online advertising [Wei07].

2.4.3.1. Purpose

Only by collecting user data and connecting this data, users' preferences can be understood and hence additional value can be delivered on a website or other Internet applications [Rub08].

2.4.3.2. Collected data

For online profiling all kind of data is collected. Mainly, there are the following sources of data: (webserver) logfiles, information collected via input provided on websites and special applications such as web bugs, cookies or software agents [Wie02]. Both descriptive data (such as customers' preferences) and identifying data (such as a customers'

⁷An IP-address is a unique number on the Internet to uniquely identify a network device such as a computer

name or address) can be collected. [Ste07] describes an experiment where webserver logfiles were analyzed to extract a click-stream⁸ of users to evaluate the most interesting websites for these users.

2.4.3.3. Threats to privacy

Threats to privacy arise when this collected data is used to create user profiles and (for example) search engine results, advertisements or products are displayed because of individualized profiles without the knowledge of users [Miy08, Xu07, Moo05] or digital dossiers are created by collecting and connecting user data available [Gro05]. Another example would be amazon.com's tests how certain "variables affect customers' purchasing decisions" by offering the same product for a different price - depending on the customer and the browser used [Ros00]. The above introduced experiment is another example how powerful such methodologies are and what threats to privacy arise when data about users is collected and connected.

All this introduced technologies demonstrate that the trade-off between privacy and services on the Internet (such as personalization) is a much discussed issue [Awa06]. However, it is also shown that users are more willing to provide personal data for personalized services than for personalized advertisement [Awa06]. Unfortunately both applications use the same technologies and methodologies to provide such personalized services. Therefore it sometimes can be hard to distinguish if a cookie is "good" or "bad" or if profiling for personalized services does make sense or not, especially for a computer. That is why new technologies and standards are necessary to provide an answer for the following question: which data is collected by a certain website and what is it used for? That is exactly the issue where privacy standards such as P3P come into play and that is why the following section is going to provide an overview about common privacy standards.

⁸A click-stream is a sequence of websites users have been visiting. It is collected to analyze the way they surf on the website, the duration of their visit, which websites are more popular than others and to analyze where visitors came from

3. Privacy Standards

The following chapter deals with widely accepted privacy standards. Although dozens of such specifications were published, this thesis will focus on proposed standards by big organizations (such as IBM) or norming institutions (such as the World Wide Web Consortium). The discussed standards are the Platform for Privacy Preference Project (P3P), the Enterprise Privacy Authorization Language (EPAL) and the eXtensible Access Control Markup Language (XACML). As this thesis mainly deals with P3P, the emphasis will lie on the Platform for Privacy Preference Project.

3.1. The Platform for Privacy Preference Project (P3P)

This section is going to introduce the Platform for Privacy Preference Project (P3P) based on the specification of the W3C [W3Ca] and a book written by one of the members of the project's specification process [Cra02]. Besides providing an overview about the goals and history of P3P, the author will take a closer look on the functionalities of P3P and the technologies and methodologies used. As the standard proposed by the W3C is rather extensive, this section will not discuss the standard in-depth enough to fulfill the need of knowledge for actual implementers of P3P. However, interested users should find the introduction useful to gain an overview about P3P.

3.1.1. P3P - an introduction

The Platform for Privacy Preference Project originated from the “Platform for Internet Content Selection” (PICS), was officially launched as a W3C project in May 1997 and became a recommendation in 2002 [Cra02]. The P3P specification version 1.1 was officially published in 2006 [W3Ca] whereas the World Wide Web Consortium describes P3P the following way [W3Ca]:

The Platform for Privacy Preferences Project (P3P) enables Web sites to express their privacy practices in a standard format that can be retrieved automatically and interpreted easily by user agents. P3P user agents will allow users to be informed of site practices (in both machine- and human-readable formats) and to automate decision-making based on these practices when appropriate. Thus users need not read the privacy policies at every site they visit.

In addition to that, the specification also highlights the features of P3P and what it is based upon [W3Ca]:

The P3P1.1 specification defines the syntax and semantics of P3P privacy policies, and the mechanisms for associating policies with Web resources. P3P policies consist of statements made using the P3P vocabulary for expressing privacy practices. P3P policies also reference elements of the P3P base data schema – a standard set of data elements that all P3P user agents should be aware of. The P3P specification includes a mechanism for defining new data elements and data sets, and a simple mechanism that allows for extensions to the P3P vocabulary.

P3P has two goals: First, it wants to empower websites to provide their privacy policy in a standardized, machine-readable way. Second, it wants users to understand which data is collected by websites and in what different ways this data will be used. More importantly, users should be able to decide whether they want to disclose specific data or not. That is, websites should be able to provide information on collected data and reasons why they actually collect that data - but in a way that these reasons can be automatically matched to the users' preferences.

To make this happen, the P3P specification defines ways and methodologies to perform this task. Two important factors are policy reference files and P3P policies. These policies describe websites' privacy policies in a machine-readable way. This can be done by using predefined vocabulary to describe privacy policies in place (via P3P policies using the P3P XML schema), using a predefined schema for data a website may wish to collect (the P3P Base Data Schema) and by applying methods to associate privacy policies to websites, cookies and other content (by using policy reference files).

3.1.2. Requirements

As every other standard, P3P specifies some requirements which need to be fulfilled to successfully implement P3P on a website and in user-agents. P3P is based on XML-files, therefore, the basic requirement is that all P3P related XML-files must be well-formed¹. Besides this requirement one of the most important specifications is the way the location of policy reference files is indicated. This is due to the fact that user-agents depend on the policy reference file to be able find and parse the P3P policies of websites and hence compare them with users' privacy settings. According to the P3P specification there are four mechanisms to indicate the location of a privacy policy reference file:

1. "Well-known location" method
2. HTTP header
3. HTML link tag
4. XHTML link tag

¹W3C provides a P3P validator to validate P3P policies which can be found at <http://www.w3.org/P3P/validator.html>. However, as of the time of writing, the validator was not operational

The well-known location method (which usage is strongly recommended by the W3C) refers to the mechanism of making the privacy reference file available on the website at the path “/w3c/p3p.xml”. This allows user-agents to easily find the necessary information. Additionally, by using this method websites make sure that user-agents can access the policies before any other requests of the website were transmitted which is essential for the “safe zone” requirement (see below). This method is also very useful for big websites with several hosts because it is an easy way to make sure that every host can publish its P3P policies independently.

Another mechanism is the HTTP header method which indicates the location by adding a new response header (the *P3P response*) to the HTTP header answering a request. The P3P specification defines that with the *policyref*-directive, an URI must be provided which indicates the location of the proper XML-files. An example HTTP response header to a GET request may look like this:

```
HTTP/1.1 200 OK
P3P: policyref="http://catalog.example.com/P3P/PolicyReferences.xml"
Content-Type: text/html
Content-Length: 7413
Server: CC-Galaxy/1.3.18
```

Listing 3.1: HTTP response header with P3P (Source: [W3Ca])

However, to successfully deploy this method administrators have to edit the server configuration. An alternative which does not require any change in the webserver’s configuration is the link tag method. Here, a simple (X)HTML link tag must be added to every website to indicate the P3P version and the URI where the necessary policy reference file can be found. A possible example is shown in Figure 3.2.

```
<link rel="P3Pv1"
      href="http://catalog.example.com/P3P/PolicyReferences.xml">
```

Listing 3.2: Link tag method for indicating policy reference file (Source: [W3Ca])

Besides requirements specifically applicable to P3P policies or policy reference files, also other requirements were defined:

- **Non-ambiguity:** Unless in exceptional cases, websites must be cautious not to declare multiple (non-expired) policies for a given URI. In such cases, all policies must be complied with because websites cannot make sure which policy has been fetched by user-agents. The P3P specification also defines which policy is applicable in the case of conflicting policies.

- **Multiple languages:** Although P3P is automatically processed XML, there are some human-readable elements which can be provided in different languages. In this case, the HTTP *Content-Language* header can be used to indicate the language of the policy. It is also defined that if a browser requests a certain *Accept-Language*, the server should provide the corresponding localized policy if available and that all localized versions of a policy must have the same meaning in each language. In addition to that, within P3P XML-files the *xml:lang* attribute may be used to indicate the language of the human-readable elements.
- **The "Safe Zone":** The P3P specification defines a "safe zone" which aims at fetching P3P policies (especially via the well-known location method) but with transferring only minimal, non-identifiable data. That means that user-agents should not send HTTP referrer headers to the safe zone and should not accept cookies from safe zone requests. In addition to that, a list of requirements for servers regarding the safe zone is also defined.
- **Policy and Policy Reference File Processing by User Agents:** Besides accepting only well-formed XML files, user-agents should only render P3P policies and policy reference files that conform to the P3P XML schema and they should not rely on parts of P3P policies or policy reference files that do not conform with this schema. In addition to that, user-agents must not modify the XML-files in any way to make them conform with the XML schema.
- **Security of Policy Transport:** P3P policies and policy references files should not contain sensitive information. Hence, these files do not have to be provided via HTTPS if the normal session is also non-encrypted.
- **Policy Updates:** If a website changes its P3P policy, it must apply the old policies to the data which was collected while the old policies were in place. Websites are in charge of keeping records of old P3P policies and policy reference files (including the date when they were valid) and to make sure that these policies are applied to previously collected data. If websites want to apply a new P3P policy to previously collected data, users must accept this beforehand.
- **Absence of Policy Reference File:** If there is no P3P policy available for a website, user-agents must assume that there is an empty policy reference file available which is valid for 24 hours. Therefore, they must check for a new version at least every 24 hours. However, website operators may publish a policy reference file that indicates that there is no policy available with an expire-date of more than 24 hours.
- **Asynchronous Evaluation:** User-agents do not have to fetch and evaluate P3P policies before any other HTTP transactions can take place. However, until the policy has been evaluated, user-agents should treat the website as if no policy is available.

- **User Preferences:** User-agents must provide a way to import and process preferences and should provide a way to export these preferences.

There are more requirements defined in the P3P specification but these will not be dealt with in-depth at this point. Important requirements will be highlighted accordingly in the following sections if necessary. For a full overview of P3P's requirements, the P3P specification should be consulted.

3.1.3. P3P policy reference files

Policy reference files are, as well as P3P policies, based on XML. To inform user-agents where they can find P3P policies, webmasters have to create a policy reference file. As the name already suggests, this file references existing P3P policies. By creating a policy reference file, webmasters can also define which policy is valid for which areas of a website and for how long the policy reference files or policies are valid. Additionally, website content such as images, pictures or cookies can also be associated with a certain P3P policy by making the proper statements in the policy reference file. This also applies to third-party content, that is if a website is, for example, including images from a third party. Besides being able to incorporate multiple policies (P3P policies can be either outsourced in a separate XML-file or be contained within a policy reference file), policy reference files can also declare in which language the policies are delivered.

An example policy reference file is shown in Listing 3.3. The example defines the **expiry date** relatively, that is in two days (or 172800 seconds) after the policy reference file has been downloaded by the user-agent (it is also possible to define a fixed date as an expire date by providing a time in GMT). It also defines three policies (namely policy "first", "second" and "third") which can be found in the policy-file at "/P3P/Policies.xml". The first policy is valid for the entire website except content in the directories "/catalog", "/cgi-bin" and "/servlet". The second policy is defined to be valid only for content within "/catalog" and the third policy only within "/cgi-bin" and "/servlet" but not in "/servlet/unknown".

```
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
  <POLICY-REFERENCES>
    <EXPIRY max-age="172800" />
    <POLICY-REF about="/P3P/Policies.xml#first">
      <INCLUDE /*</INCLUDE>
      <EXCLUDE /catalog /*</EXCLUDE>
      <EXCLUDE /cgi-bin /*</EXCLUDE>
      <EXCLUDE /servlet /*</EXCLUDE>
    </POLICY-REF>
    <POLICY-REF about="/P3P/Policies.xml#second">
      <INCLUDE /catalog /*</INCLUDE>
    </POLICY-REF>
    <POLICY-REF about="/P3P/Policies.xml#third">
```

```

<INCLUDE>/cgi-bin/*</INCLUDE>
<INCLUDE>/servlet/*</INCLUDE>
<EXCLUDE>/servlet/unknown</EXCLUDE>
</POLICY-REF>
</POLICY-REFERENCES>
</META>

```

Listing 3.3: Policy reference file (Source: [W3Ca])

Please note that both P3P policies and policy reference files make extensive use of **wildcards** by using the asteriks (“*”). Wildcards can be applied to every URL, for example to define that policy “fourth” is only applied to JPEG files on the website:

```

<POLICY-REF about="/P3P/Policies.xml#fourth">
  <INCLUDE>/*.jpg</INCLUDE>
</POLICY-REF>

```

Listing 3.4: Policy reference file using wildcards

Both elements INCLUDE and EXCLUDE (and their content, respectively) are case-sensitive, that is, depending on the webserver, one has to make sure that all possible file-extension combinations are accounted for.

When it comes to **third-party content**, referencing a policy is not easily accomplished as URL’s always have to be relative to the policy reference file of the website and the P3P specification does not allow the application of policies on third-party content. However, a methodology was introduced to *hint* user-agents at the policy which can be applied for third party content by using the HINT element. User-agents should first check the well-known location of the third-party website for applicable P3P policies. If no (applicable) policies can be found, user-agents may evaluate the HINT element to be directed to the third-parties’ P3P policies. However, the HINT element may only be used if the third-party site declares the location of its policy reference file either via a HTTP header or LINK tag (because the well-known location was already checked before the HINT element was evaluated). The HINT element can be used by defining a *scope* of the policy and a *path* relative to the website where the policy reference file is located as shown in Listing 3.5

```

<HINT scope="http://www.example.org" path="/mypolicy/p3.xml" />
<HINT scope="http://www.example.net:81" path="/w3c/prf.xml" />
<HINT scope="http://*.shop.example.com" path="/w3c/prf.xml" />

```

Listing 3.5: Policy reference file using the HINT element (Source: [W3Ca])

Another important issue for policy reference files arise when it comes to **cookies**. For the association of policies with cookies, P3P specifies the two elements `COOKIE-INCLUDE` and `COOKIE-EXCLUDE`. For both elements the *name* of the cookie, the *value*, *domain* and *path* of the cookie has to be provided. The example shown in Listing 3.6 defines that the policy “first” applies to all cookies except for the one with the name “obnoxious-cookie” from “.example.com” and that actually the second policy is applied to this cookies from domain “.example.com”.

```
<META xmlns="http://www.w3.org/2002/01/P3Pv1">
  <POLICY-REFERENCES>
    <POLICY-REF about="/P3P/Policies.xml#first">
      <COOKIE-INCLUDE name="*" value="*" domain="*" path="*" />
      <COOKIE-EXCLUDE name="obnoxious-cookie" value="*"
        domain=".example.com" path="/" />
    </POLICY-REF>
    <POLICY-REF about="/P3P/Policies.xml#second">
      <COOKIE-INCLUDE name="obnoxious-cookie" value="*"
        domain=".example.com" path="/" />
    </POLICY-REF>
  </POLICY-REFERENCES>
</META>
```

Listing 3.6: Policy reference file using the `COOKIE-INCLUDE` and `COOKIE-EXCLUDE` elements (Source: [W3Ca])

In addition to the already introduced features, policies can also be applied to certain **HTTP methods** using policy reference files. That means that in a policy reference file, a certain policy can be applied to one of the HTTP methods such as `OPTIONS`, `GET`, `HEAD`, `POST`, `PUT`, `DELETE`, `TRACE` and `CONNECT` (cp. [W3Cb] for more details on the HTTP/1.1 specification). Practically applicable in most cases are probably the `POST` and `GET` methods to apply policies to either requested data or sent content via forms. The corresponding element to be used for this purpose is `METHOD`.

P3P offers a diverse set of features for policy reference files of which the most important ones were introduced here. For a more detailed description, readers should consult the P3P specification available at [W3Ca].

3.1.4. P3P policies

P3P policies are the heart of P3P. With a P3P policy, a website can transfer its privacy policy in a machine-readable, standardized way which enables user-agents to parse this policy and compare it to user-preferences in browsers, proxies or other software. To be standardized and machine-readable, P3P policies have to use a predefined vocabulary and syntax. Within a P3P policy, there are five general assertions which apply to the

entire policy and six specific assertions (which are used in *statements*) which only apply to the collected data. The five general assertions are:

- With the two attributes *discur* and *opturi* of the POLICY element the location of the human-readable privacy policy (also called the *disclosure URL*) and possible *opt-out* mechanisms (if applicable) have to be provided.
- The TEST element indicated whether a P3P policy is for testing purposes only. This is especially useful if websites are implementing P3P and want to test their P3P configuration. User-agents should not consider policies with the TEST indicator set as valid. This element is optional.
- The ENTITY element contains all relevant contact information about the website operators, the *entity* running the website. This includes name, address, e-mail, telephone and other contact data.
- Websites may allow users to access, modify or delete the data the website has collected about them. For the purpose of indicating the level of access, the ACCESS element was introduced.
- Another optional element concerns the dispute resolution. If website visitors have a privacy-related dispute, they should know whom to contact and how to contact the corresponding party. This information can be provided using the DISPUTES element. If using the DISPUTES element, websites should also include the sub-element REMEDIES to specify which remedies are applicable if websites do not adhere to their privacy policy.

Listing 3.7 shows the general assertions of microsoft.com’s P3P policy. It also shows that the privacy policy of microsoft.com is available at <http://go.microsoft.com/?linkid=4412894> and that registered users can “view, add, or edit the personal information and marketing contact preferences” Microsoft stores about users at Microsoft’s Profile Center which is available at <https://profile.microsoft.com/RegSysProfileCenter/Infodefault.aspx>. Beside Microsoft’s contact information, it is also stated that Microsoft provides access to all identified data they collected (`<ACCESS><ALL/></ACCESS>`) and that there are two ways to resolve dispute issues: either by contacting Microsoft Customer Service or TRUSTe whereas both will correct any errors made but no monetary compensation will be provided (`<REMEDIES><CORRECT/></REMEDIES>`).

```
<POLICY xmlns="http://www.w3.org/2000/12/P3Pv1"
  discuri="http://go.microsoft.com/?linkid=4412894"
  opturi="https://profile.microsoft.com/RegSysProfileCenter/
  Infodefault.aspx">
<ENTITY>
  <DATA-GROUP>
    <DATA ref="#business.name">Microsoft Corporation</DATA>
```

```

<DATA ref="#business.contact-info.postal.street">1 Microsoft Way<
  /DATA>
<DATA ref="#business.contact-info.postal.city">Redmond</DATA>
<DATA ref="#business.contact-info.postal.stateprov">WA</DATA>
<DATA ref="#business.contact-info.postal.country">USA</DATA>
<DATA ref="#business.contact-info.postal.postalcode">98052-6399</
  DATA>
<DATA ref="#business.contact-info.online.email">
  homepage@microsoft.com</DATA>
<DATA ref="#business.contact-info.online.uri">http://support.
  microsoft.com/contactus/?ws=mscom</DATA>
</DATA-GROUP>
</ENTITY>
<ACCESS<all /></ACCESS>
<DISPUTES-GROUP>
<DISPUTES
  resolution-type="service"
  service="http://support.microsoft.com/contactus/?ws=mscom"
  short-description="Microsoft_Customer_Service">
  <LONG-DESCRIPTION>If for some reason you believe microsoft.
    com has not adhered to these principles, please notify us
    by e-mail at homepage@microsoft.com</LONG-DESCRIPTION>
<REMEDIES<correct /></REMEDIES>
</DISPUTES>
<DISPUTES
  resolution-type="independent"
  service="http://www.truste.org/users/watchdog.html"
  verification="Truste"
  short-description="TRUSTe_Certification">
  <LONG-DESCRIPTION>Microsoft is a premier sponsor of TRUSTe and a
    member of the TRUSTe privacy program, an independent, non-
    profit initiative whose mission is to build users' trust and
    confidence in the Internet by promoting TRUSTe's principles of
    fair information practices.</LONG-DESCRIPTION>
  <IMG src="http://www.microsoft.com/library/images/gifs/profilectr/
    Truste.gif" width="91" height="73" alt="TRUSTe: Click to Verify
    " />
<REMEDIES<correct /></REMEDIES>
</DISPUTES>
</DISPUTES-GROUP>

```

Listing 3.7: microsoft.com's P3P policy - general assertions

Besides general assertions, there are also six data specific assertions which are encapsulated by a STATEMENT element. This enables websites to group data elements they collect in a statement. If a website for example collects user-data in the logfiles of their

webservers for administrative purposes as well as data to process customer orders, it might have two statements. These data specific assertions are:

- A *consequence* should provide the users with information specifying why their data is actually collected and why it may be valuable for a certain service, transaction or feature. It is stated in a human-readable form using the CONSEQUENCE element. This element is optional, its usage however is strongly recommended.
- If a websites does not collect identifiable data (so called *non-identifiable data*) or if it anonymizes this data, it may set an indicator accordingly using the NON-IDENTIFIABLE element. However, websites should note that the P3P specification provides a strict definition of *anonymized*: 'In order to consider the data "anonymized", there must be no reasonable way for the entity or a third party to attach the collected data to the identity of a natural person' [W3Ca]. This element is optional.
- How the data a website collects is going to be used has to be provided using the PURPOSE element. The P3P vocabulary defines twelve purposes whereas the twelfth provides a possibility to provide human-readable information. The other eleven purposes have a predefined meaning. Table 3.1 recites the "Plain Language Translations of P3P Vocabulary Elements" of the P3P specification for these twelve elements according to [W3Ca].
- The RECIPIENT element states with which parties collected data will be shared. P3P specifies six types of recipients which are listed in Table 3.2.
- The data-retention policy in effect must be defined using the RETENTION element. Although no specific time is indicated, the five sub-elements can give users an indication about the websites' retention policy which can be supported by a human-readable retention policy. Table 3.3 lists all sub-elements including their meaning.

As with policy reference files, P3P policies can be assigned a **language** by using the *xml:lang* attribute. This attribute is used to identify the language of human-readable fields.

Another feature already highlighted is the policy **lifetime**. By setting the EXPIRE element, website can provide information on how long a certain policy is valid, that is when user-agents have to re-fetch a policy for a website or certain content elements.

Besides "normal" P3P policies, one can also define **compact policies** (CP) which represent a summary of a websites' P3P policy for a cookie. Compact policies are transmitted via an additional HTTP response header CP. As the name suggests, compact policies do not provide the same details as P3P policies. However, compact policies are heavily used especially by Microsoft's Internet Explorer which decides if a cookie is going to be blocked or not solely based on compact policies [Cra02].

PURPOSE	Plain Language Translation
current	To provide the service you requested.
admin	To perform web site and system administration.
develop	For research and development, but without connecting any information to you.
tailoring	To customize the site for your current visit only.
pseudo-analysis	To do research and analysis in which your information may be linked to an ID code but not to your personal identity.
pseudo-decision	To make decisions that directly affect you without identifying you, for example to display content or ads based on links you clicked on previously.
individual-analysis	To do research and analysis that uses information about you.
individual-decision	To make decisions that directly affect you using information about you, for example to recommend products or services based on your previous purchases.
contact	To contact you through means other than telephone (for example, email or postal mail) to market services or products.
historical	To aid in historical preservation as governed by a law or policy described in this privacy policy.
telemarketing	To contact you by telephone to market services or products.
other-purpose	For other uses: [include site's human, readable explanation; if site omits human-readable explanation say "not described here"].

Table 3.1.: PURPOSE sub-element definitions and translations (Source: [W3Ca])

RECIPIENT	Plain Language Translation
ours	Companies that help us fulfill your requests (for example, shipping a product to you), but these companies must not use your information for any other purpose.
delivery	Delivery companies that help us fulfill your requests and who may also use your information in other ways.
same	Companies that have privacy policies similar to ours.
other-recipient	Companies that are accountable to us, though their privacy policies may be different from ours.
unrelated	Other companies whose privacy policies are unknown to us.
public	People who may access your information from a public area, such as a bulletin board, chat room, or directory.

Table 3.2.: RECIPIENT sub-element definitions and translations (Source: [W3Ca])

RETENTION	Plain Language Translation
no-retention	We do not keep your information beyond your current online session.
stated-purpose	We keep your information only long enough to perform the activity for which we collected it.
legal-requirement	We keep your information only as long as we need to for legal purposes.
business-practices	Our full privacy policy explains how long we keep your information.
indefinitely	We may keep your information indefinitely.

Table 3.3.: RETENTION sub-element definitions and translations (Source: [W3Ca])

Finally, an example should be provided to explain the STATEMENT element and hence the data-specific assertions. Listing 3.8 shows how a statement in a P3P policy may look like when a form is provided to e-mail webmasters comments about the website including an input field to optionally provide an e-mail address so that the webmaster can reply to the inquiry: the statement defines that the data submitted will only be used to improve the website, that the information will be discarded once processed and that the information provided will not be disclosed to any other party.

```

<STATEMENT>
  <PURPOSE><current /><admin /></PURPOSE>
  <RECIPIENT><ours /></RECIPIENT>
  <RETENTION><stated-purpose /></RETENTION>
  <DATA-GROUP>
    <DATA ref=' '#dynamic.miscdata ' '>
      <CATEGORIES>
        <content />
      </CATEGORIES>
    </DATA>
    <DATA ref=' '#dynamic.miscdata ' ' optional=' 'yes ' '>
      <CATEGORIES>
        <online />
      </CATEGORIES>
    </DATA>
  </DATA-GROUP>
</STATEMENT>

```

Listing 3.8: Policy statement regarding a e-mail comment form (Source: [Cra02])

As seen in the example above, this statements makes use of the DATA and CATEGORIES elements. These elements can be used within a policies' DATA-GROUP element. Within such a DATA-GROUP element, data types which the policy is applied to have to be defined whereas such data types can be grouped in *data schemas*. P3P specifies a *base data schema* where the following kind of data is defined: dynamic data, user data, third party data and business data. As the base data schema has a multitude of sub-(sub)-elements, it is not going to be covered in this thesis. Interested readers should especially consult chapter five of the P3P specification for more details. A complete P3P policy with all elements introduced can be found in Appendix A.1 for further reference.

So far, predefined vocabularies were introduced which were designed by the P3P working group. Although a lot of effort and time was devoted in creating the vocabulary as short as possible and as long as necessary, it cannot contain every possible concept. That is

why **extensions** were developed. By using the EXTENSION element new concepts can be tested and implemented. This element is very flexible and can be placed almost in all P3P specified elements. Interested readers should consult the P3P specification for more details.

3.1.5. A P3P Preference Exchange Language (APPEL)

Although P3P specifies how websites can express their privacy policies, it does not offer a solution for expressing user preferences can be expressed. This is done by a separate W3C specification called APPEL - A P3P Preference Exchange Language. The goal of APPEL (pronounced a-pell) is to save and exchange users' preferences (which APPEL refers to as *rule-sets*) in a standardized way. The reason for this is that most P3P user-agents probably won't show all possible P3P settings because there would be too much combinations to be handled by end-users. However, predefined, sensible privacy settings should be made available to and distributed by users and other parties so that end-users are able to easily decide what level of privacy they want to apply. APPEL-files are XML-files and contain *rule-sets* which include a pattern to be matched against a P3P policy and an action which is to be executed when a match is found. Readers interested in this topic should consult the W3C Working Draft on APPEL for more information [W3Cc]. It should also be mentioned that the two biggest P3P user-agents (namely Microsoft's Internet Explorer and AT&T's Privacy Bird) do not (fully) support APPEL.

3.1.6. Existing P3P user agents and software tools

For users interested in P3P, there are currently two noteworthy implementations: P3P in Microsoft's Internet Explorer and AT&T's Privacy Bird² which is a plugin for Microsoft Internet Explorer 5.01, 5.5, and 6.0. The Mozilla Suite once supported P3P but P3P support was removed with bug-report 225287 [Moz03]. Neither Mozilla Firefox nor Seamonkey supported P3P. There was an extension available for Firefox but it is not compatible anymore with current versions of Firefox and can be designated as a proof-of-concept implementation only (see the Part II of this thesis for more details).

For implementors of P3P and website operators there are several tools available which support them in their endeavour:

- The website www.p3ptoolbox.org provides a useful implementation guide
- The European Commission Joint Research Centre published the JRC Policy Workbench which is a suite to edit and test P3P policies. The project is available at SourceForge³.
- IBM published its IBM P3P Policy Editor which is available at⁴

²<http://www.privacybird.org>

³<http://sourceforge.net/projects/jrc-policy-api>

⁴<http://alphaworks.ibm.com/tech/p3peditor>

- There is a commercial P3P policy generator available at⁵
- When it comes to the validation of P3P policies, there is a W3C validator available at⁶

The P3PToolbox also mentions other tools which are not listed here⁷ as well as the W3C P3P website⁸. Although P3P 1.1 is compatible with P3P 1.0, it should be highlighted that some of the tools mentioned have not been updated in the last few years and therefore do not necessarily generate strict P3P 1.1 compatible XML.

3.1.7. Future of P3P

The future of the P3P project is uncertain. At the homepage of the project itself it is written that “there was insufficient support from current Browser implementers for the implementation of P3P 1.1” [W3Ce]. That is the reason why the work on P3P 1.1 has been suspended and P3P 1.1 was not published as a recommendation but rather as a working group note. W3C’s statement that it “is not excluded that W3C will push P3P 1.1 until Recommendation if there is sufficient support for implementation” [W3Ce] sound rather half-heartedly when considering the next entry on the project’s homepage: A new group (called “PLING” - Policy Languages Interest Group) has been created “to discuss interoperability, requirements and related needs for integrating and computing the results when different policy languages [are] used together” [W3Ce]. It can be questioned whether a group which discusses the interoperability of standards which have never been broadly accepted and implemented will create additional value.

3.2. The Enterprise Privacy Authorization Language (EPAL)

The Enterprise Privacy Authorization Language was developed by International Business Machines Corporation (IBM) and is described by IBM as “a formal language for writing enterprise privacy policies to govern data handling practices in IT systems according to fine-grained positive and negative authorization rights. It concentrates on the core privacy authorization while abstracting data models and user-authentication from all deployment details such as data model or user-authentication” [IBM03]. Unless mentioned otherwise, the following section about EPAL is based on IBM’s EPAL 1.2 specification available at [IBM03].

⁵<http://p3pedit.com/>

⁶<http://www.w3.org/P3P/validator/20010928/>

⁷<http://www.p3ptoolbox.org/tools/resources1.shtml>

⁸<http://www.w3.org/P3P/implementations.html>

3.2.1. EPAL - an introduction

EPAL was developed by IBM and is, similar to P3P, based on XML-files. The two main goals of EPAL are to “provide the ability to encode an enterprise’s privacy-related data-handling policies and practices” and to be a “language that can be imported and enforced by a privacy-enforcement systems” [IBM03]. As these two statements suggest, EPAL has a different goal than P3P. Whereas P3P wants to make **published** privacy policies on enterprises’ websites machine-readable with predefined vocabularies, “EPAL aims at formalizing **enterprise-internal** privacy policies” [IBM03]. This requires a very detailed set of vocabulary specifically tailored to the needs of enterprises. Unlike P3P, EPAL has the specific goal not to be a language with a universal vocabulary. As long as all involved parties agree on and understand the vocabularies being used, organizations can define a set of vocabularies which fits their needs. In addition to that, EPAL policies can be enforced as they are parsed by an enforcement engine. That means that EPAL was ‘developed mainly as a business-to-business (B2B) technology to help streamline information flows during business interactions’ and that “it helps [to] ensure that information is protected and used in accordance with the responsible organizations privacy policies” [Stu04]. However, EPAL has, similar to P3P, some requirements which are going to be dealt with in the next section.

3.2.2. Requirements

The EPAL specification defines requirements which are not designed for any specific platform or software. It rather aims at defining requirements how EPAL policies and their meaning actually have to be formed and expressed: One requirement is that a rule may be “allowed” or “denied”, have no defaulting ruling at all, or that each rule must contain a set of conditions that must be true to make the rule active. It is also a requirement that a rule may contain obligations which have to be executed if an action has to be performed (e.g. that accessing a certain file must be documented in a specific way).

3.2.3. EPAL policies

Unlike P3P, EPAL does not have any predefined elements. “EPAL provides a mechanism for defining the elements which are used to build the policy” [IBM03]. To build an EPAL policy, a set of rules has to be compiled whereas these rules have to be ordered with descending precedence. Similar to P3P, EPAL has elements which define a rule statement: a ruling, a user category, an action, a data category, and a purpose - and it optionally can contain conditions and obligations. The following example shows how to convert a privacy policy into an EPAL privacy rule. The informal privacy policy

*Allow a sales agent or a sales supervisor to collect a customer’s data for order entry if the customer is older than 13 years of age and the customer has been notified of the privacy policy. Delete the data 3 years from now.
(Source: [IBM03])*

Element	Value
ruling	allow
user category	sales department
action	store
data	category customer-record
purpose	order-processing
condition	the customer is older than 13 years of age
obligation	delete the data 3 years from now

Table 3.4.: An example EPAL privacy rule (Source: [IBM03])

Element	Value
user category	sales department
action	store
data	category customer-record
purpose	order-processing

Table 3.5.: An example EPAL request (Source: [IBM03])

would be translated into the formal EPAL privacy rule as shown in Table 3.4. Such rules are used to determine whether a request is allowed or not. Not surprisingly, a request contains a user category, an action, a data category, and a purpose. Assume the following informal request:

A person acting as a sales agent and an employee requests to collect a customer's email for order entry. (Source: [IBM03])

This informal request would be translated into the formal EPAL request as shown in Table 3.5, based on the elements predefined by EPAL and used by the above created privacy rule. This request matches the rule defined above and therefore it would be allowed.

As EPAL was especially designed for enterprises, it offers vocabularies that enables businesses to express *sector-specific privacy policies*. Although usually one enterprise will create a policy, the aim is also that companies agree on a set of vocabularies, exchange these policies and hence use them in their business transactions. The following section is going to highlight these EPAL vocabularies.

3.2.4. EPAL vocabularies

By using the EPAL-VOCABULARY element, industry-specific vocabularies can be defined using the following subelements:

- VOCABULARY-INFORMATION: This element provides information about the vocabulary - the name (*id*), the issuer of the vocabulary (*issuer*) and the version (*version-info*).

- **USER-CATEGORY:** This element represent categories of users (such as a single employee, particular roles, departments etc.). User-categories are defined and then used in a *rule* to describe who is going to access the data.
- **DATA-CATEGORY:** This element classifies data in different categories - medical records for example would be in another category than an e-mail address of a customer. Similar to **USER-CATEGORY**, this element is referenced in *rules*.
- **PURPOSE:** This elements defines the purpose for what the data is used. However, this purpose may be high-level (e.g. marketing), therefore it is needed to build hierarchies to represent different kinds of purpose (e.g. telemarketing vs. third-party e-mail marketing).
- **ACTION:** The ultimate goal for collecting data is to process it somehow. This element describes the actions that are going to be applied to the collected data. A privacy policy will then define which actions are allowed or denied under specific circumstances.
- **CONTAINER:** This element contains context data in a structured form (e.g. attribute *age*) so that these attributes can be evaluated by *conditions* (e.g. if age is over 13 then...).
- **OBLIGATION:** Obligations are actions which have to be taken after performing an action on data, for example that some data must be deleted within 30 days.

The detailed (XML-) structure of EPAL and its policies and rules is not very different from the concepts already introduced and will not be covered here. Interested readers should consult the EPAL specification available at [IBM03] for more in-depth information.

3.3. The eXtensible Access Control Markup Language (XACML)

The eXtensible Access Control Markup Language (XACML) is a standard published by OASIS, the Organization for the Advancement of Structured Information Standards, and defines an access control language based on XML [OAS05]. “The motivation behind XACML is to express the well-established ideas in the field of access-control policies (e.g., rules, policies, policy sets, subjects, decision requests, authorization decisions) using an extension language of XML” [Cov08a]. The XACML specification states that it was developed because “there is a pressing need for a common language for expressing security policy [sic]. If implemented throughout an enterprise, a common policy language allows the enterprise to manage the enforcement of all the elements of its security policy in all the components of its information systems” [OAS05].

If not stated otherwise, the following introduction is based on the XACML 2.0 specification available at [OAS05].

3.3.1. XACML - an introduction

Similar to P3P and EPAL, XACML is implemented using XML-files. It also uses the already common approach of defining one *policy* which includes several *rules*. It is checked whether a rule applies and what the effect of this rule is (the request is allowed or denied) based on given *conditions*. Such matches are being performed if a *request* has been sent whereas the necessary information to decide whether the request matches the rule under given conditions is provided with the request. These similarities are no coincidence as EPAL 1.1 used XACML explicitly and EPAL 1.2 uses a lot of XACML's concepts such as attributes, rules, conditions, etc [Sun05]. Furthermore, XACML is based on a previous project of IBM - the XML Access Control Language (XACL) [Cov08b].

However, [Sun05] highlights some difference between EPAL and XACML, especially when it comes to the evaluation of rules. Whereas in EPAL the first rule which conditions are met is applicable (the descending precedence of rules was already described above), XACML has a combined algorithm in place. That means if in XACML policies several rules match one request, a choice is given to the user which rule to apply. Although there are some other key differences between EPAL and XACML, EPAL remains a subset of XACML - interested readers should consult [Sun05] for more detail on that issue.

3.3.2. Requirements and structure of XACML

The XACML specification defines some non-normative requirements which are generic requirements for a policy language. It does, however, of course also defines XACML specific requirements which are going to be shortly discussed in this section:

- The three basic XACML elements are POLICYSET, POLICY and RULE. Similar to the previously introduced privacy standards, a policy-set can contain one or more policies and one policy can contain one or more rules. Data access requests have to go through a *Policy Enforcement Point* (PEP). This PEP formulates a request for an authorization decision which is then sent to a *Policy Decision Point* (PDP). This PDP evaluates the existing policies and sends the authorization decision (*Permit*, *Deny*, *NotApplicable*) back to the PEP.
- To decide on the authorization, the combining algorithm mentioned above comes into place. It has to ensure that only one policy or policy-set is applicable. If no policy or policy-set is applicable, it returns *NotApplicable*, if more than one policy or policy-set is applicable, it returns *Indeterminate*.
- XACML supports *Role Based Access Control* (RBAC). This means that a *subject* can hold a *role* (e.g. IT-Manager) and that such roles have to be considered when making authorization decisions.

This collection of specifications was just exemplary. For a full list of requirements readers should consult the XACML specification.

Once all the requirements have been fulfilled, requests can be matched to policies. Listing 3.9 shows a basic example of an XACML request. It states that the *subject*, Bart Simpson with the e-mail address bs@simpsons.com, wants to access (to be precise *read*) a certain *resource*, in this case his medical file at path file://example/med/record/patient/Bart-Simpson. This request would then be matched with the corresponding policy and if applicable, the request would be allowed or denied.

```
<?xml version="1.0" encoding="UTF-8" ?>
<Request xmlns="urn:oasis:names:tc:xacml:2.0:context:schema:os"
  xmlns:xsi="http://www.w3.org/2001/XMLSchema-instance"
  xsi:schemaLocation="urn:oasis:names:tc:xacml:2.0
:context:schema:os http://docs.oasis-open.org/xacml/
access_control-xacml-2.0-context-schema-os.xsd">
  <Subject>
    <Attribute AttributeId="urn:oasis:names:tc:xacml:1.0
      :subject:subject-id" DataType="urn:oasis:names:tc:xacml:1
      .0:data-type:rfc822Name">
      <AttributeValue>
        bs@simpsons.com
      </AttributeValue>
    </Attribute>
  </Subject>
  <Resource>
    <Attribute AttributeId="urn:oasis:names:tc:xacml:1.0
      :resource:resource-id" DataType="http://www.w3.org/2001/
      XMLSchema#anyURI">
      <AttributeValue>
        file://example/med/record/patient/BartSimpson
      </AttributeValue>
    </Attribute>
  </Resource>
  <Action>
    <Attribute AttributeId="urn:oasis:names:tc:xacml:1.0
      :action:action-id" DataType="http://www.w3.org
      /2001/XMLSchema#string">
      <AttributeValue>
        read
      </AttributeValue>
    </Attribute>
  </Action>
</Environment />
</Request>
```

Listing 3.9: Basic XACML request example (Source: [OAS05])

3.3.3. Summary

The introduced privacy standards all fulfill different needs on different (technical) levels and perspectives. Whereas P3P is more consumer orientated and tries to inspire companies to make their privacy policy easily accessible via user-agents, EPAL takes a look at internal privacy needs of businesses. It helps businesses to make sure privacy is being enforced by giving them a way to express their internal privacy policies and to set-up industry-specific vocabularies. This empowers companies to share such a set of vocabularies which in return enables enterprises to enforce privacy policies during business to business (B2B) transactions. XACML is a standard which deals with privacy-issues on a lower level. Therefore it is, if not a basis, then at least a model for other privacy standards.

4. Requirements for Privacy Tools

In this chapter, software quality requirements for privacy tools will be introduced. These defined requirements provide the basis for Chapter 5 where existing privacy tools will be evaluated according to these characteristics. The requirements for privacy tools are based on the following standards:

- ISO/IEC 25030:2007 which “provides requirements and recommendations for the specification of software quality requirements” [ISO07]
- Part one of ISO/IEC 9126:2001 which defines a quality model describing characteristics of internal and external quality and quality in use [ISO01a]
- Part four of ISO/IEC 9126:2001 which defines “metrics for measuring attributes of quality in use defined in ISO/IEC 9126-1:2001” [ISO01b]
- IEEE Std 830-199 Recommended Practice for Software Requirements Specifications which (among other things) aims at providing requirements which can be applied when selecting software [IEE98]

Most of these standards were also addressed in [Boe08]. Although all standards have certain measurements in place to evaluate specific requirements or attributes (especially ISO/IEC 9126-4:2001), a textual description of the requirements as provided in Chapter 5 will suffice for the purpose of this thesis. However, this section will distinguish between generic software quality requirements (based on the above mentioned standards) and privacy specific requirements, especially the support of privacy standards and the prevention of threats.

4.1. Generic software quality requirements

There has been a lot of research and standardization focused on software quality requirements as the above mentioned standards demonstrate. Although different software fulfills all kind of needs, some basic quality requirements remain the same. That is why generic software quality requirements - which also apply for privacy software - are going to be discussed in the following section.

4.1.1. Functionality

ISO/IEC 9126:2001 defines functionality as “the capability of the software product to provide functions which meet stated and implied needs when the software is used under

specified conditions” [ISO01a]. That means that functionality deals with fulfilling specific needs for users whereas other characteristics are concerned with how and when it fulfills these needs [ISO01a]. An example would be if users install net-banking software on their computer - what does the software do to fulfill the users’ needs? IEEE Std 830-199 defines more fine-grained functional requirements which should be considered when creating a software requirement specification [IEE98].

4.1.2. Reliability

This characteristic is probably one of the most obvious one for users. As wear and tear should not apply to software, reliability is defined as “the capability of the software product to maintain a specified level of performance when used under specified conditions” [ISO01a]. If a software is not able to maintain a specific level of performance the fault mostly lays in the implementation or design of the software, for example if a cache is not emptied and the software is therefore operating slower than usual [ISO01a]. Such factors are reflected in the maturity of a software and should be considered when specifying the requirements of a software [IEE98]. Reliability also includes fault tolerance and recoverability, for example after a certain error has happened.

4.1.3. Usability

Usability is defined as “the capability of the software product to be understood, learned, used and attractive to the user, when used under specified conditions” [ISO01a]. Although some areas of functionality and reliability also effect usability, these are not classified as usability in ISO/IEC 9126:2001. Among other things usability also incorporates understandability, attractiveness and learnability of software.

4.1.4. Efficiency

ISO/IEC 9126:2001 defines efficiency as “the capability of the software product to provide appropriate performance, relative to the amount of resources used, under stated conditions” [ISO01a]. That is that under given resources, software should behave in a timely way and resources should be allocated well. Part four of ISO/IEC 9126:2001 defines certain metrics to measure efficiency (such as task effectiveness, task completion or error frequency) [ISO01b].

4.1.5. Maintainability

Maintainability is “the capability of the software product to be modified” whereas “modifications may include corrections, improvements or adaptation of the software to changes in environment, and in requirements and functional specifications” [ISO01a]. This characteristic is, obviously, focused on the maintenance of the software itself [IEE98] and hence includes factors such as the changeability, stability or testability of a software product.

4.1.6. Portability

ISO/IEC 9126:2001 defines portability as “the capability of the software product to be transferred from one environment to another” [ISO01a]. This characteristic aims at highlighting how flexible software is regarding a change of environment whereas this could be a change of organisational, hardware or software environment [ISO01a]. Therefore, this requirement is aimed at the question of how adaptable and replaceable a software is and how it can co-exist with other software ([ISO01a], [IEE98]).

4.1.7. Quality in use

From a user perspective, this requirement is the most important one. Quality in use is defined as “the capability of the software product to enable specified users to achieve specified goals with effectiveness, productivity, safety and satisfaction in specified contexts of use” [ISO01a]. This characteristic is measured not from the properties of the software itself but rather from the results of using the software. Part four of ISO/IEC 9126:2001 defines metrics for all four specified goals: effectiveness, productivity, safety and satisfaction [ISO01b].

4.2. Privacy specific requirements

Despite general software quality requirements there are privacy specific requirements which are especially important when it comes to the evaluation of privacy tools. As these privacy specific requirements were not found in the literature studied by the author, the following privacy specific requirements were defined because they reflect the additional requirements needed to evaluate privacy specific software, in the author’s opinion.

4.2.1. Support of privacy standards

This characteristic defines how privacy tools support privacy standards such as P3P, EPAL or XACML. Of course it is also of interest which other standards are supported, if any. It is also considered whether certain standards are partially or fully implemented in a specific software. Although this characteristic is defined as a privacy specific requirement, it is also important to evaluate how the above mentioned general software quality requirements are incorporated: If a certain privacy standard is shipped with a software, does it meet usability criteria? Does it provide all functionalities which were defined by the standard? How reliable is the software?

4.2.2. Prevention of threats

When it comes to software which supports privacy, the prevention of threats is an important feature. Therefore the following questions should be answered: which kind of threats does the software help to ban? Which abilities does the software have to prevent such threats? Again, the above defined general software quality requirements are to be

applied to this characteristic too, especially functionality, usability and reliability. In addition to that, the prevention of threats also includes that privacy software considers the exposure to potentially harmful technologies such as JavaScript or Flash. These technologies are of course not a harmful technology per se but can be abused for example to install Malware¹ via malicious websites [FSec]. That means that privacy software can also be directed at the prevention of such an exposures.

Using these defined characteristics, existing tools (such as privacy tools) can be evaluated. Although this list of requirements does not satisfy the level of detail necessary for developing software or creating a detailed software requirements specification, it is suffice to evaluate a set of privacy tools as done in Chapter 5. In addition to that, it provides a potential basis for future work on requirements specification for privacy software tools.

¹Malware is malicious software such as trojan horses or viruses

5. Evaluation of Existing Privacy Tools

In this chapter, existing privacy tool will be evaluated and on overview about the most popular representatives of browsers, plug-ins and proxy-software in regards to privacy will be given. These three technologies were chosen because browsers are the ultimate interface for users to access the Internet and by using plug-ins and proxy-software, users are able to reduce privacy risks accordingly. Although all evaluations are based on the criteria defined in Chapter 4, they are not scientifically measured and are solely based on a textual description. However, to ensure that all tested tools have the same technical basis, the evaluation will be executed in a sandbox (that is Windows XP SP3 on Microsoft Virtual PC 2007). As most of the tools have a different objective, it is not always possible to compare them.

5.1. Browsers

In this section, the most frequently used browsers [W3S09] will be evaluated in terms of build-in P3P support: Microsoft's Internet Explorer, Mozilla's Firefox, Apple's Safari and Opera Software's Opera browser.

5.1.1. Microsoft Internet Explorer

Since version 6 Microsoft's Internet Explorer (MSIE) has built in support for P3P Version 1 compact policies [MS07a, MS07b]. Although users may not be aware of it, the "Privacy" settings in the "Internet Options" of MSIE define how to treat cookies by using a "Zone-System": In the "Local Intranet" and "Trusted" zones, all cookies are accepted. When using the "Medium" level (which is set by default), MSIE "blocks third-party cookies that do not have a compact policy [...] or third-party cookies that have a compact policy that specifies that personally identifiable information is used without your implicit consent." [MS07a]. First-party cookies that store personally identifiable information without implicit consent are downgraded to session cookies whereas first-party cookies without a compact policy are restricted so that they can be read only by the issuing domain. In the "Restricted" zone, all cookies are blocked [MS07a].

When visiting a website which does not match the users' privacy preferences, a Privacy Report icon is displayed in the MSIE status bar which users can double click to view the websites' privacy report [MS07c]. Although the privacy report clearly parses P3P policies and shows the generated information to the user via the privacy report (cp. Figure 5.2), the user does not have any options to actively change settings regarding P3P as MSIE only deals with compact policies. There is a way to import so called

“Customized Privacy Settings” using XML-files but still, one can only define how MSIE handles cookies [MSDN] - there is no way to configure any other settings P3P offers.

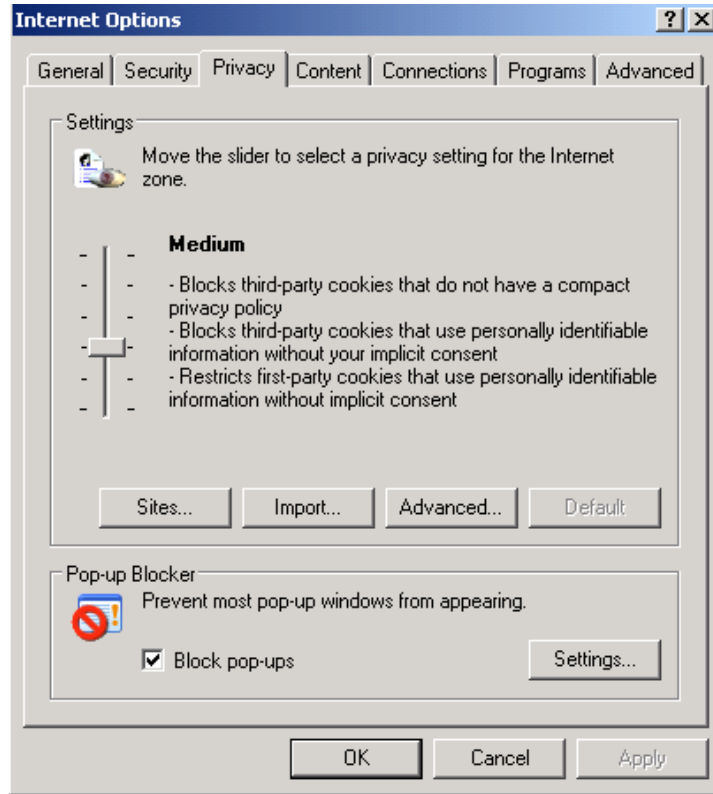


Figure 5.1.: Microsoft Internet Explorer 6.0 privacy settings

At the time of writing, Release Candidate 1 (RC1) of Microsoft’s Internet Explorer version 8.0 has been released. However, no detailed information on any changes regarding P3P could be found nor does MSIE 8.0 handle P3P enabled website differently although this behaviour might change in future releases.

However, there is a plugin called “Privacy Bird” for MSIE 5.01-6.0 available which adds P3P support to the browser. The software was originally developed by AT&T and was then further developed by Carnegie Mellon University. Unfortunately, there has not been an update to Privacy Bird in the last few years so the software is out of date and cannot be used with current versions of Microsoft’s Internet Explorer anymore. Furthermore, it does not work unless the users has administrator rights and the last operating system supported was Windows XP. The authors made the source code available at their website but until yet, no other software has been developed based on Privacy Bird.

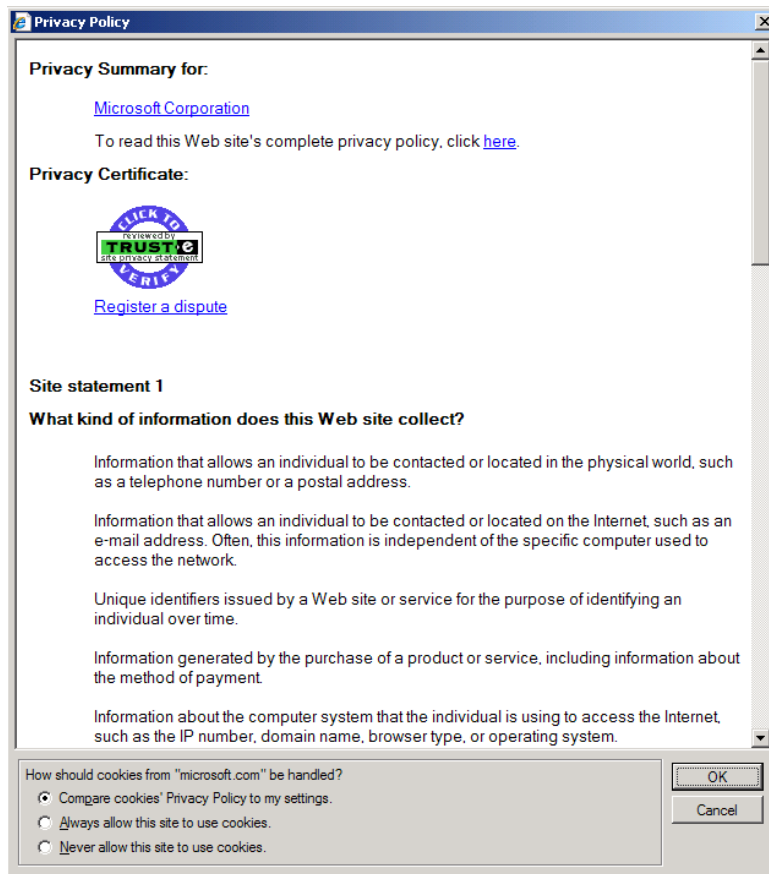


Figure 5.2.: Microsoft Internet Explorer 7.0 privacy report on www.microsoft.com

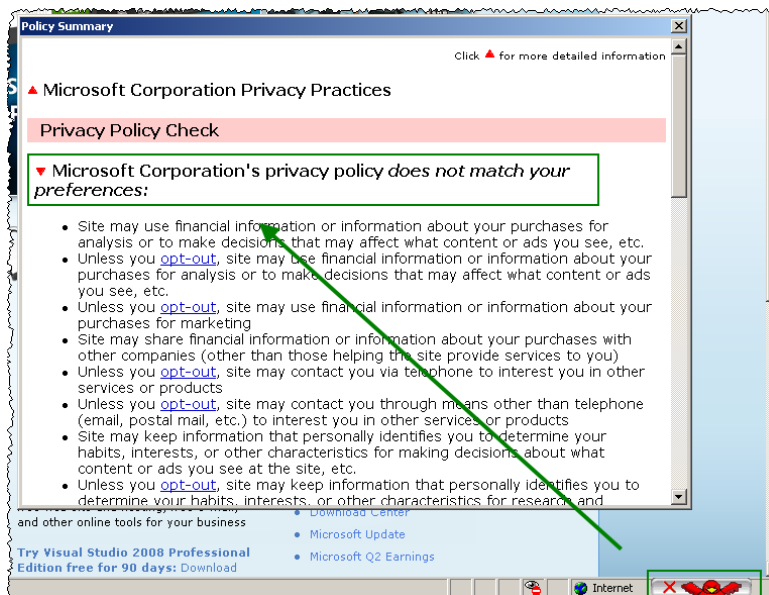


Figure 5.3.: Privacy Bird for Microsoft Internet Explorer

5.1.2. Firefox

The Mozilla Foundation had built in P3P support in its software. But after a long discussion it was eventually removed completely from its browser-line and so neither SeaMonkey nor Firefox have built in P3P support anymore [Moz07]. One of the reasons for the removal was that the code “has not been owned or touched since 2003” [Moz07]. Michael Kaply from IBM even stated: “Ah the memories. We (IBM) wrote the original P3P implementation and then Netscape proceeded to write their own. So both our companies wasted immense amounts of time that everyone thought was a crappy proposal to begin with. Remove it.” [Moz03]. However, Firefox 3 now by default blocks third party cookies.

5.1.3. Safari

Apples’ Safari for Windows does not offer any privacy settings whatsoever. No information is displayed when third-party cookies are blocked, no privacy report is shown or could be accessed. The only option users can set is a radio-box whether they want to accept third-party cookies or not as shown in Figure 5.4.

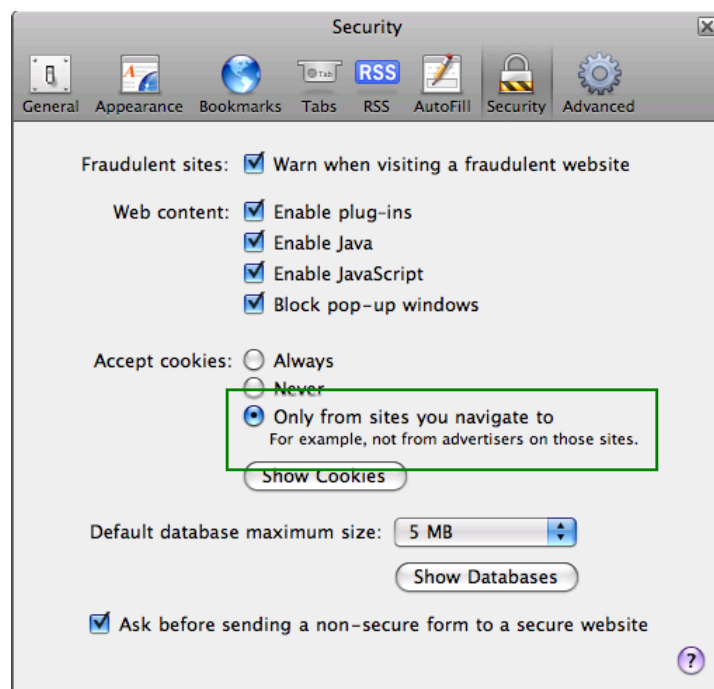


Figure 5.4.: Apples’ Safari security settings

5.1.4. Opera

Similar to Safari, Opera does not offer any privacy settings except blocking third-party cookies (cp. Figure 5.5). Blocked cookies are also not shown when browsing the Internet. In an article published at InformationWeek.com in 2001, Live Leer, at that time PR

manager for Opera Software AS, stated that “at the moment, we aren’t sure whether P3P is the best solution” [IW01] - which may explain the lack of Opera’s P3P support.

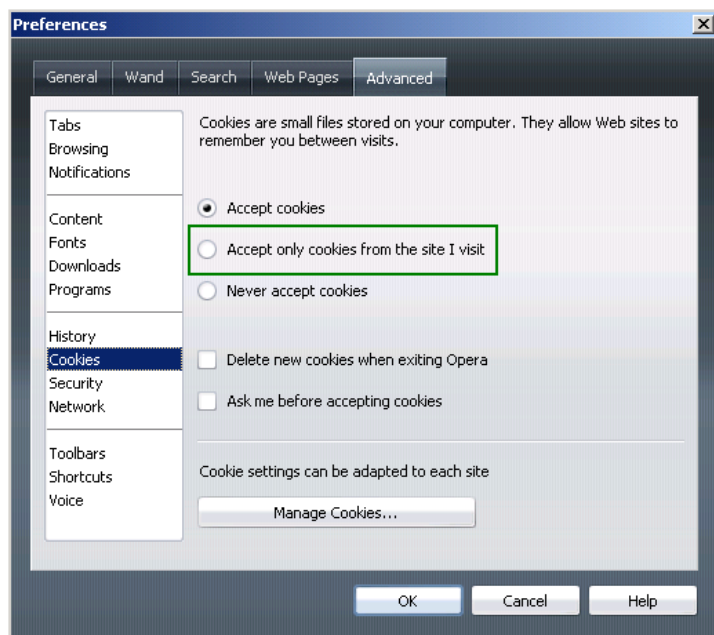


Figure 5.5.: Opera’s cookie settings

5.2. Plug-ins

The next section will briefly introduce selected plug-ins (or extensions) for Mozilla Firefox to enhance privacy. Due to the huge amounts of extensions available only some of the most popular ones could be picked for evaluation.

5.2.1. Adblock Plus

Adblock Plus is a Firefox extension that claims to eliminate 99% of advertisements shown on the web. The principle is simple: If you visit a website, the extension blocks all advertisements it can recognize. To recognize as many ads as possible, users provide extensive lists which include adservers, typical URLs of ads and even CSS-code which is normally used to position or display ads. This extensions has its use especially on websites which are overloaded with ads such as MySpace.com.

5.2.2. NoScript

NoScript allows the execution of active content such as JavaScript, Java and other plug-ins (Flash, Silverlight etc.) only if the user has defined the website as a trusted site. Therefore it is ideal to protect users from known or unknown scripting attacks such as cross-site scripting (XSS). Although it (temporarily) allows the execution of scripts via

shortcuts and context-menu, users should take some time to initially configure NoScript as most websites depend on JavaScript and/or Java to work properly. However, in the authors' view NoScript is only suited for advanced users as unexperienced users may not have the knowledge to configure the tool correctly and hence may be frustrated that their favourite websites do not work anymore.

5.2.3. Flashblock

As the name suggests, Flashblock prohibits the display of flash-content unless the website is listed in the white-list. It activates the flash-content by clicking on it as shown in Figure 5.6.

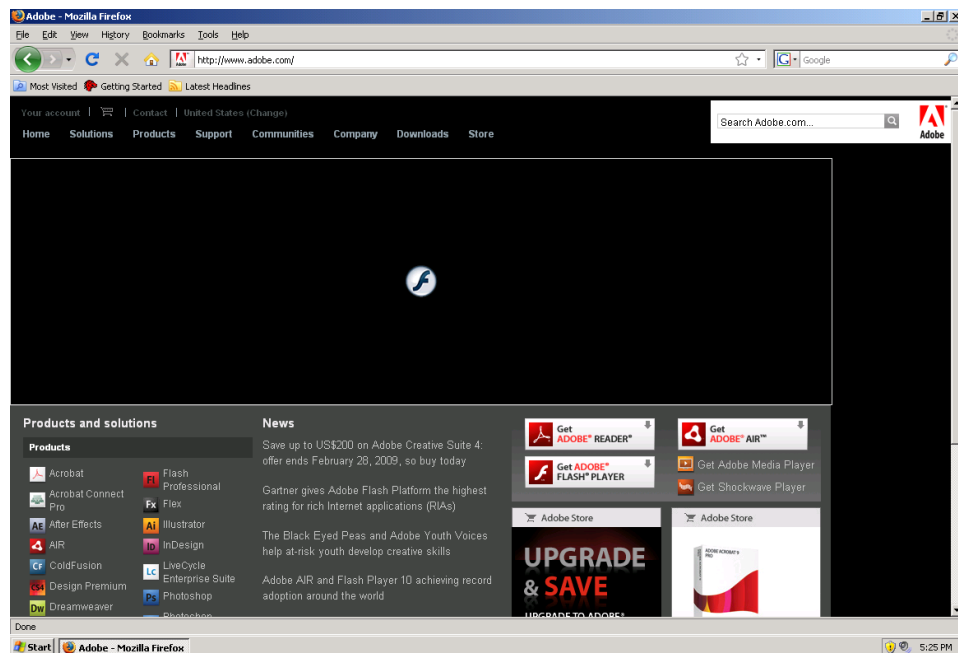


Figure 5.6.: Flashblock in action: Press the “Flash”-button to start the flash animation

5.2.4. BugMeNot

With BugMeNot users can bypass compulsory web registration by using the content menu of Firefox and logging in with the data provided via www.bugmenot.com. This helps users not to provide any sensitive information to third parties by using already existing logins as demonstrated for www.nytimes.com as shown in Figure 5.7. Users can rate the success rate of logins on www.bugmenot.com and submit new logins.

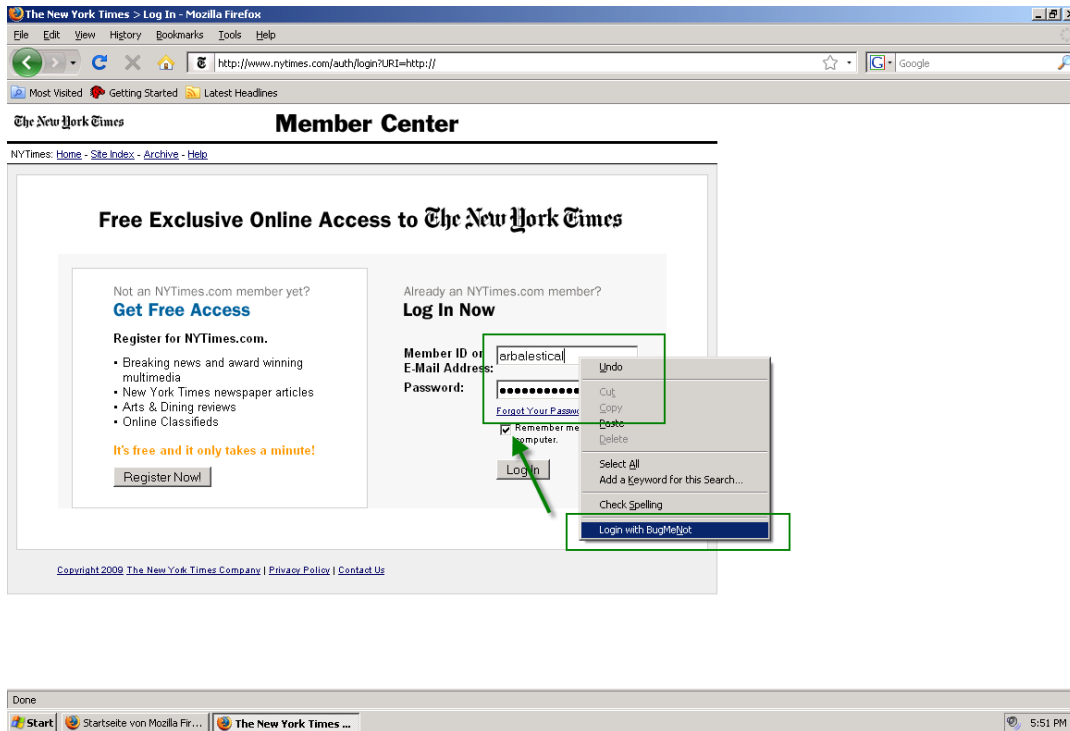


Figure 5.7.: Logging in onto www.nytimes.com by using BugMeNot

5.2.5. TrackMeNot

TrackMeNot tries to protect users against profiling of search engines by randomly fire off queries to search engines in the background while surfing the Internet. The tool not just generates a lot of traffic in the background, it can also be questioned if it actually works as search engines probably have ways to differ automatically from manually submitted queries. If users want to make sure that their search queries are totally anonymized, they should use an extension such as Private Web Search or a proxy as described next.

5.2.6. Private Web Search (PWS)

Private Web Search (PWS) is a Firefox extension that acts as a proxy to anonymize users' search queries. Technically seen it acts as an HTTP proxy which filters the HTTP request. It then sends the search query to the search engine via the Tor anonymity network. The HTTP Proxy receives the answer from the Tor network and then returns the results back to PWS. The result is a normal answer to a search query but no sensitive (e.g. personal identifiable) information should have been transferred to the search engine. [Sai07] describes PWS and the technical background in more detail. Unfortunately the extension could not be evaluated due to difficulties to install and setup PWS.

5.3. Proxies

In this section, proxies will be discussed which can be used to enhance privacy while surfing the Internet. A proxy, or proxy server, “is a computer system or router that breaks the connection between sender and receiver” [ZDN09a]. In this case anonymous proxies are of special interest as they hide “the IP address of the user’s machine from the Web site and may provide encryption on the user side” [ZDN09b].

5.3.1. Anonymizer

An anonymizer is a tool which routes all user-generated Internet traffic via anonymous proxies. The aim is to leave as few traces behind as possible while surfing the Internet. There are many tools available to achieve this goal, one of the most popular ones is “Java Anon Proxy” (JAP) which is available at <http://anon.inf.tu-dresden.de>. Originally a scientific project of a German university, it is now also commercially available as “JonDonym”. The principle of the system is easily explained: By redirecting the traffic over a mix of servers which commit to not storing any logfiles, users’ traffic can be partly anonymized. JAP can be downloaded as a Java application (cp. Figure 5.8) and there is also a Firefox extension available (which is in its alpha status though).

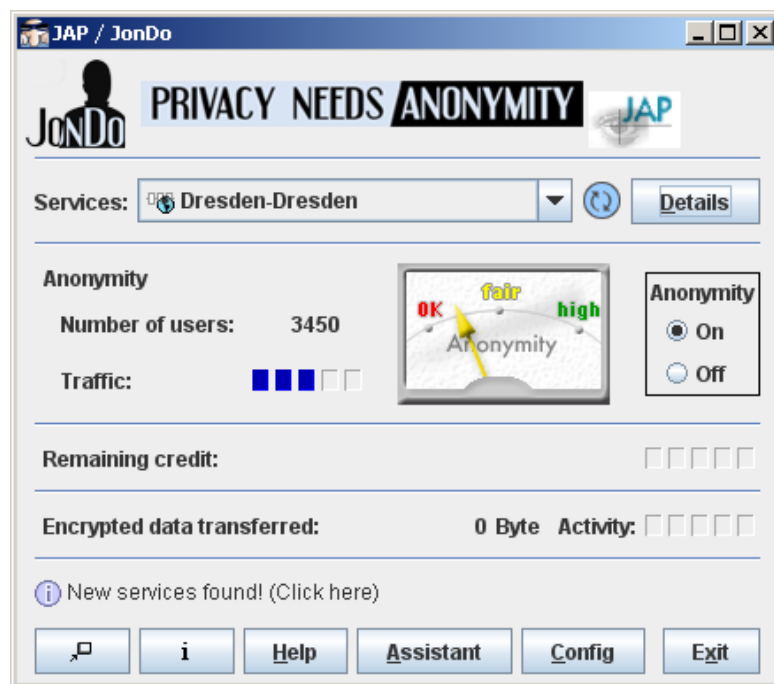


Figure 5.8.: Java Anon Proxy user-interface

5.3.2. The Onion Router (TOR)

The Onion Router (“Tor”) is a network of several nodes which enable anonymous communication over the Internet by encrypting the traffic and sending the traffic via an

indirect route to its destination. The nodes in the network are called “onion routers” because they uncover a layer of the “onion” (the traffic-package) to receive instructions where the traffic should be relayed to next - at the end, the package arrives at its final destination in plain text [OR09]. The Tor software makes use of this principle and can be obtained directly at the Tor project website at <http://www.torproject.org>. There are ready to install bundles available at the website of the project. These bundles also install all software necessary to use Tor with Firefox (including the “Torbutton” extension). With the “Tor Map” tool one can see all onion routers (that is nodes) in the Tor network and which route the traffic took for example while browsing “www.google.com” as shown in Figure 5.9. Although the Tor network may provide a huge amount of anonymity, its disadvantage clearly is the slow speed of the network due to a lack of bandwidth. While conducting tests, the bandwidth available decreased from approx. 9MB/s without Tor to approx. 0-14kb/s (!) with Tor which makes normal surfing impossible.

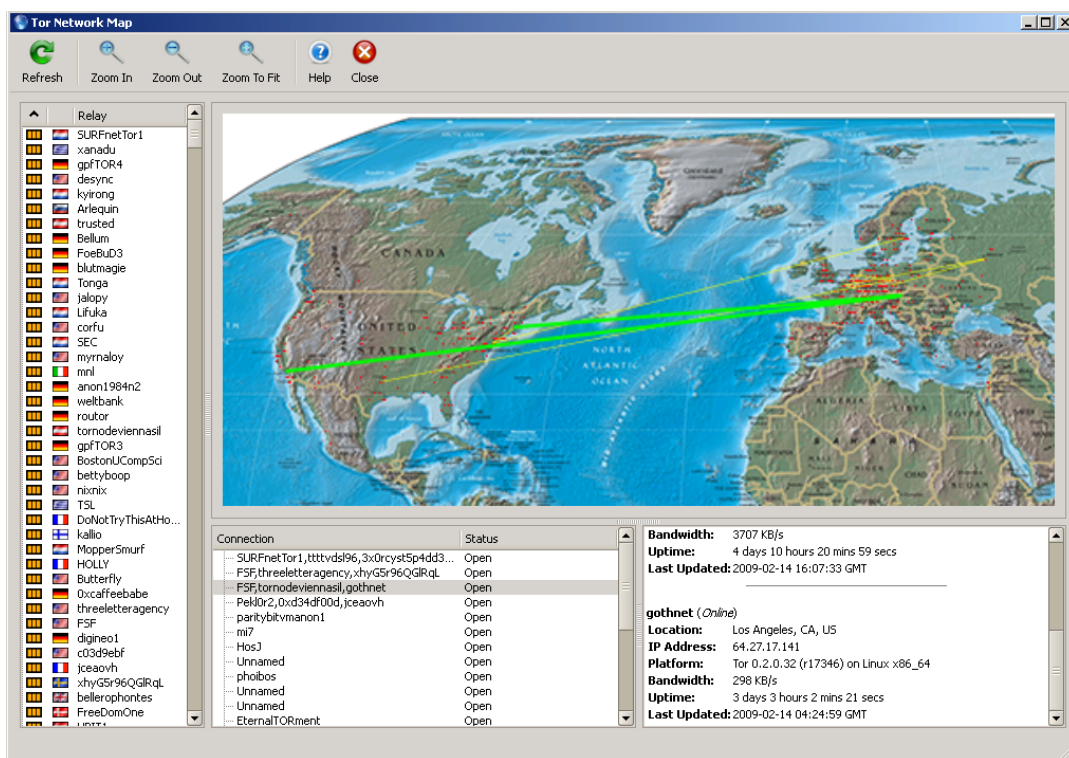


Figure 5.9.: Map of onion routers and the route between two nodes used to access “www.google.com”

5.3.3. Switchproxy

Besides these sophisticated solutions, there are of course also simple Firefox extensions for managing different proxies. One example is Switchproxy which easily enable users to switch proxy with two mouse clicks, for example to change from the ISPs’ proxy to the proxy of ones’ university to access the digital library without having to change the connection settings from Firefox manually.

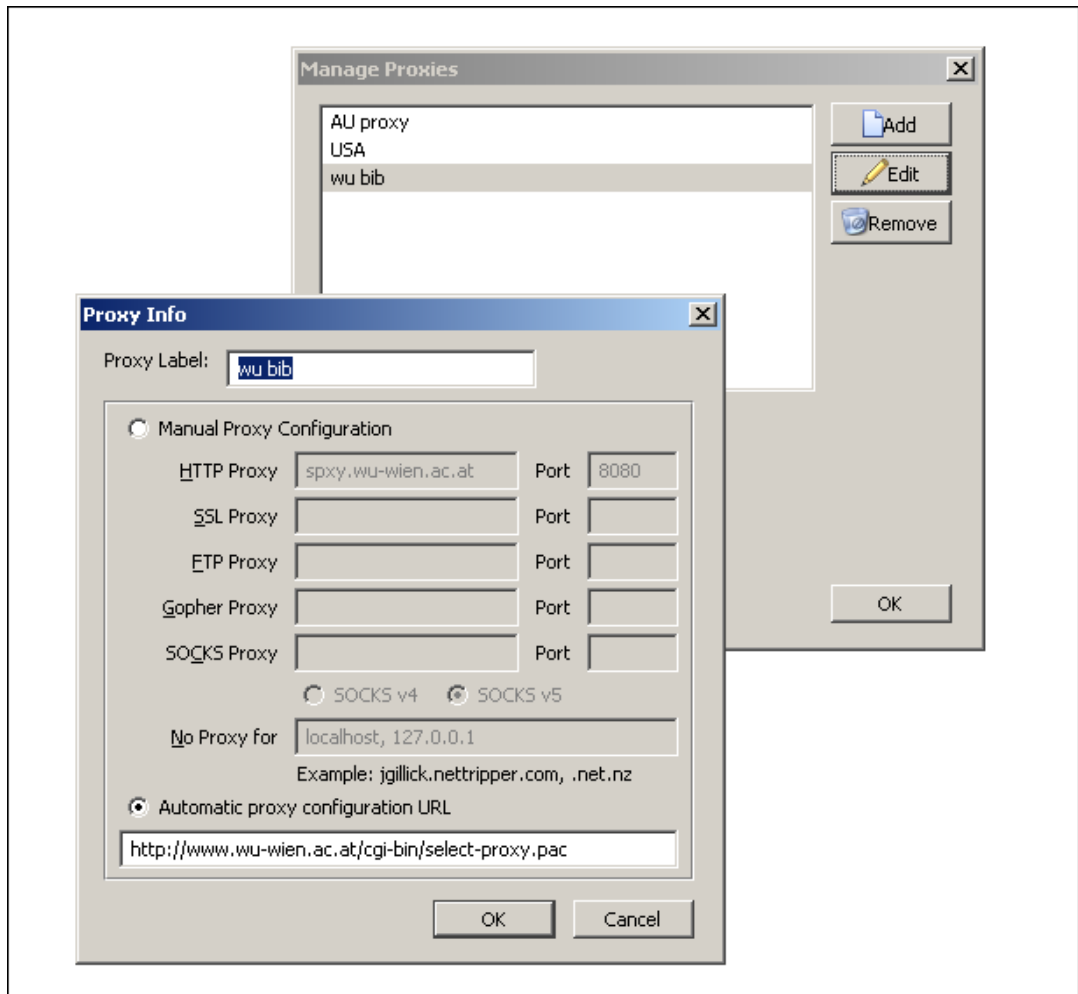


Figure 5.10.: The settings dialog of Switchproxy

Part II.
Applied part

6. Development of a Privacy Plug-In

In this chapter, the applied part of this master thesis will be discussed. As already mentioned earlier, the applied part consists of the development of a Mozilla Firefox extension or to be precise: an enhancement and up-port of an outdated P3P extension for Firefox to create a basis for the open source community to further enhance it.

Before describing the initial situation of the project, some general information on extension development for Mozilla Firefox will be provided. Afterwards, the enhancement of the existing extension will be briefly discussed. The chapter will then be finished by the evaluation of the newly developed Firefox extension.

6.1. Add-on development for Firefox

The area of add-on or extension development for Mozilla Firefox is wide and fills books [Fel07]. That is why in the upcoming part of this thesis, only a brief overview on this topic can be given. However, this work tries to provide useful resources and references if the need for in-depth information arises. Parts of this section have originally been written by the author in [EPAB09].

6.1.1. The extension concept

With its concept of extensions, Mozilla provides a flexible way to add new functionalities to its products. While this concept is applicable to all Mozilla products (such as SeaMonkey, Thunderbird, Sunbird etc), this chapter only deals with Firefox specifics which may or may not be applicable to other Mozilla products. It has to be highlighted that extensions are, at least in the Mozilla scope, something different than plug-ins. Plug-ins deal with more “core”-functionalities of Firefox such as Adobe’s PDF-reader or similar. Although both extensions and plug-ins can be based on native code (such as C++), extensions are mostly developed using JavaScript and XUL (pronounced “zool”) which is an XML dialect [MDC09a].

As just mentioned, extensions are basically based on two technologies: JavaScript and the XML User Interface Language (XUL). Whereas this thesis is not going to discuss JavaScript, XUL should shortly be introduced as it is not as well known as JavaScript. XUL is “Mozilla’s XML-based language that lets you build feature-rich cross platform applications that can run connected or disconnected from the Internet” [MDC09b]. XUL basically allows developers to easily create graphical user interfaces (GUI) with all known elements such as buttons, status-bars, menu-lists, drop-down lists, etc. by using an XML dialect. Besides drag and drop support, XUL also offers so called “XUL templates”.

With XUL templates, one is able to dynamically create XUL elements. Currently XUL templates support RDF and XML-files as data sources. SQLite support was just recently introduced with Firefox 3 [MDC09c].

6.1.2. Developing Firefox extensions

There are many tutorials on the Internet explaining how to build extensions for Firefox. One of the best found was the one available at rietta.com [RIET]. Although the suggested tutorial gives a good introduction, it does not deal with advanced topics such as flexible GUIs, XUL templates, using SQLite databases, storing and saving files on local hard discs, building extension by using Eclipse and Ant build-files, loading components, debugging extensions and so on.

That is why this section is going to briefly describe the general structure of Firefox extensions and is then hopefully going to add value by transcribing the experiences made while undertaking the applied part of this thesis. However, no working code examples will be provided at this stage as this would go beyond the scope of this thesis.

6.1.3. File structure

Firefox extensions are distributed using a zip-file as a container which is saved with an “.xpi”-file-extension (pronounced “zippy”). The contents of the xpi-file have the following structure [MDC09d]:

```
exampleExt.xpi:
/install.rdf
/components/*
/components/cmdline.js
/defaults/
/defaults/preferences/*.js
/plugins/*
/chrome.manifest
/chrome/icons/default/*
/chrome/
/chrome/content/
```

Basically there are five important directories and files to mention:

- **install.rdf**: the install.rdf-file is an Install Manifest which is an RDF/XML-styled file. The Mozilla Developer Center (MDC) defines an Install Manifest as “the file an Add-on Manager-enabled XUL application uses to determine information about an add-on as it is being installed. It contains metadata identifying the add-on, providing information about who created it, where more information can be found about it, which versions of what applications it is compatible with, how it should be updated, and so on” [MDC09e].

- **components** directory: this directory can hold XPCOM Components. These are JavaScript or C++ libraries which provide Interfaces which are registered with Firefox and can then be used by the extension. [MDC09d] describes this in more detail.
- **defaults** directory: all JavaScript files which are stored in this directory are going to be executed when the extension is installed. Therefore, it can be used to define preferences using the Mozilla Preferences API which extensions can then use. Another interface for the preferences is the (probably well known) technique of typing “about:config” into the Firefox URL-bar [MDC09f].
- **chrome.manifest**: by using the chrome.manifest-file, developers can change the Firefox GUI. They can, for example, add entries in menu bars or toolbars by “overlying” customized XUL-files. This is done by “registering” XUL-files with Chrome as described in more detail in [MDC09g].
- **chrome/content**: this is the directory where the actual JavaScript-, XUL-files, etc. are stored. There can also be directories for skins and localization if your extension has the need for it - see the Mozilla Developer Center at [MDC09h] for more information on these topics.

6.1.4. Flexible graphical user-interfaces (GUIs)

XUL offers the option of defining windows which are flexible, that is they “flex” to the resolution and space on the screen which is available. Obviously it is highly recommended not to design GUIs with a fixed sized, that is by defining height and width as the chances are very high that the GUI will not always be displayed properly. Although every XUL element can be designed by using Cascading Style Sheets (CSS), this works only if a “flex” attribute with a value greater than zero has been defined for each XUL container which is affected. By using this flexible system, one can define how empty space in a container (such as a window) should be occupied by the existing elements and (relatively to the other elements) how much of this empty space each element should occupy - depending on the flex value. The result of using the “flex-system” is, that elements of a GUI resize automatically if the resolution or window-size is changed.

6.1.5. XUL templates

By using XUL templates, developers are able to dynamically create XUL elements. As already mentioned, there are currently three types of data sources available: RDF-files, XML-files and SQLite databases. However, the principle is the same: developers can query a data source and dynamically create XUL elements such as items in a menu drop down list. See [MDC09i] for more details.

6.1.6. Accessing SQLite databases

Although the developed extension does not use SQLite databases, the topic should be briefly discussed for the sake of completeness. SQLite databases have been introduced with Firefox 3.0. They are easy to handle and can be queried using SQL. By using the mozStorage API one can directly open a connection to the database and query it by using JavaScript. The results can then be used to create XUL elements or simply by building content via JavaScript. [MDC09j] provides an example and more details.

6.1.7. Accessing the local filesystem

Once installed, an extension runs in the scope of the Firefox browser and has therefore much higher rights on the user's computer than for example websites which are visited by using the browser. That means that extensions can access (and therefore also delete) all files and directories, even recursively! [MDC09k] provides a good overview about the API and how to use it, including some working examples.

6.1.8. Using an integrated development environment (IDE)

The author strongly recommends using an IDE such as Eclipse to develop Firefox extensions. It makes the development process a lot easier, for example by using Subversion (for version management) and Ant to automatically build the xpi-files. There are tools such as XulBooster (an Eclipse Plugin)¹, the "Extension Developer's Extension" (a Firefox Extension)² or the "Firebird/Thunderbird Extension Wizard"³ available which also help in setting up a proper environment.

6.1.9. Debugging Firefox

For debugging Firefox extensions, one needs a JavaScript debugger such as Venkman⁴ which is Mozilla's official JavaScript debugger. Although it provides useful insights, it sometimes cannot help developers, for example, finding out why extensions are not installed properly or why components are not loaded. This is due to the fact that Venkman itself comes as an extension and therefore is loaded the same time as other extensions are loaded and therefore cannot interrupt start-up processes. To accomplish this, developers have to run Firefox in debug-mode and set breakpoints in the Firefox source-code itself. More information on that issue can be found at the MDC. Another very important tool for debugging Firefox is a DOM Inspector. A DOM Inspector such as "DOMi", Mozilla's DOM Inspector⁵, lets you inspect the document structure of XUL windows and the browser. It is for example very useful to find out how to address XUL elements if the documentation is not clear in that point.

¹<http://cms.xulbooster.org>

²<http://ted.mielczarek.org/code/mozilla/extensiondev/>

³<http://ted.mielczarek.org/code/mozilla/extensionwiz/>

⁴<https://developer.mozilla.org/en/Venkman>

⁵https://developer.mozilla.org/en/DOM_Inspector

Speaking of information it has to be mentioned that it sometimes can be hard to find the correct information in terms of deprecated- or up-to-date information. With hindsight one of the most useful resources for this project was the source code of Firefox itself as it is - in this regards - the best documented source of information found on the web. Mozilla also provides a tool which interested developers can use to search the source code at <http://mxr.mozilla.org/firefox>.

6.2. Initial situation

As already mentioned, the aim of the applied part of this thesis is to enhance and up-port an outdated P3P extension for Firefox to create a basis for the open source community to further enhance it. The outdated P3P extension for Firefox is called “Privacyfox” and was originally developed by Fahd Arshad [Ars04] whereas the new, enhanced version which was developed by the author of this thesis and is called “Webprivacy”. Privacyfox and the issues which were encountered during the transition from Privacyfox to Webprivacy will be described in this section.

6.2.1. Compatibility

Privacyfox was developed in 2004 and there has not been a new release since 2005. Therefore, the extension does not work with current versions of Firefox anymore. This is due to several reasons, mainly they are security related:

- Privacyfox tries to open a new tab and to write directly into it. This insecure behaviour is not allowed anymore in Firefox 3.
- Privacyfox writes its debugging messages into a tab, not to the error console provided by Firefox.
- Privacyfox loads external P3P policies directly from the Internet into the browser which is also not allowed anymore in Firefox 3.

In addition to that, Privacyfox has some limitations when it comes to P3P 1.1 compliance. The most crucial ones are:

- The “Test”-element is not validated.
- No support of multiple policies in one file.
- The “Expiry”-element is not checked upon.
- Lacking support of three mechanisms to locate policy reference files: Privacyfox only checks for a policy reference file in the well-known location, not in an HTML link-tag, XHTML link-tag or in the HTTP header.
- Privacyfox does not check cookies, images or other external content on P3P compact policies.
- Privacyfox does not check the privacy policy of websites before the website is actually loaded into the browser window.

6.2.2. Functionality

Besides compatibility issues with Firefox and the P3P standard, there are also some issues which have to be addressed regarding the functionalities of Privacyfox:

- Privacyfox can only check for P3P policies if the domain name does not contain any hyphens (e.g. the hostname “www2-ibm.com” cannot be checked).
- There is no support whatsoever for setting user-specific preferences.
- Hence, no actual check can be made whether the currently visited website actually matches the users’ privacy preferences.
- Furthermore, no visualization is available which indicates users the current P3P status of the website (e.g. if it matches or does not match the users’ preferences, if the website has a P3P policy at all, if the policy is well formed or not etc.).
- The option to display the P3P policy is hidden in the “View”-menu of Firefox. When clicking the corresponding menu entry (cp. Figure 6.1), Privacyfox simply opens a new tab and displays the parsed P3P policy as shown in Figure 6.2.

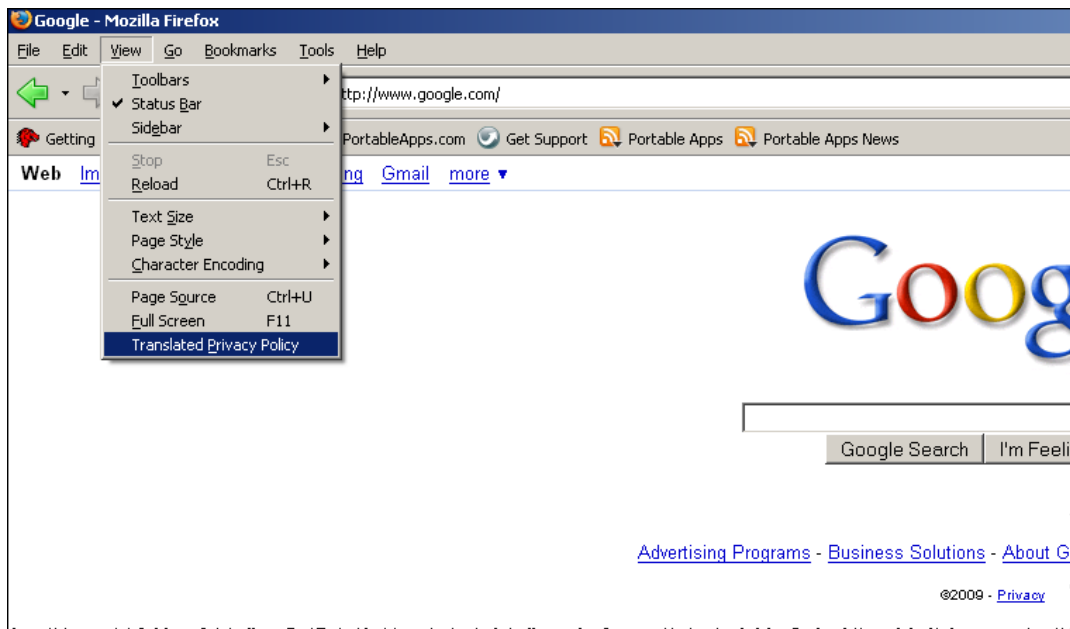


Figure 6.1.: Accessing the P3P policy translated by Privacyfox

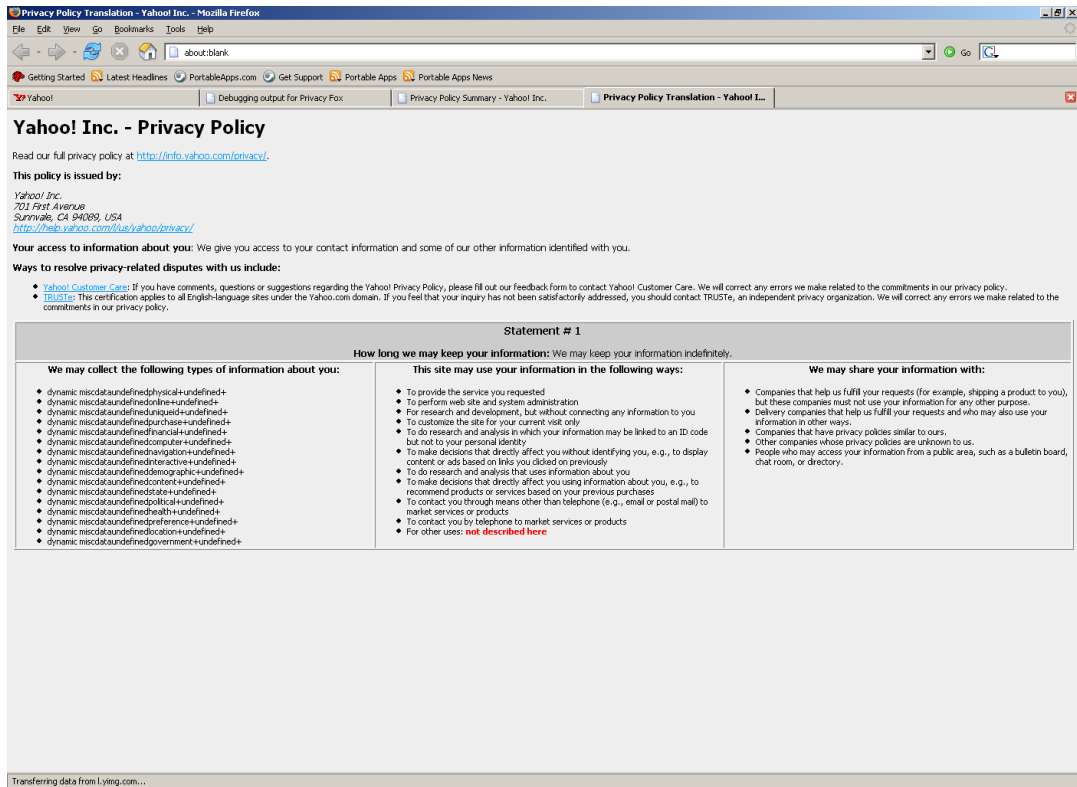


Figure 6.2.: Displaying the P3P policy translated by Privacyfox

However, it has to be mentioned that the author of Privacyfox was aware of some of these issues (cp. [Ars04, page 8ff.]). Although most of the important problems regarding functionality and usability have been resolved by Webprivacy, some of the problems regarding P3P compliance still exist with the Webprivacy extension as the next section is going to discuss in detail.

6.3. Up-port and enhancement of a P3P add-on

The initial idea for conducting the applied part of this thesis was to use Privacyfox as the basis for Webprivacy. However, after a short time it was clear that most of the code from Privacyfox had to be rewritten and that important parts of Webprivacy had to be developed from scratch.

6.3.1. Compatibility

When it comes to compatibility, the first step was to make sure that the extension could be used with Firefox 3:

- Writing the human-readable translation directly into a tab was removed and replaced by using a separate XUL-window.

- All debugging messages are also not printed into a tab anymore but are rather written into the error console of Firefox.
- Of course it is possible to enable or disable the debugging-mode and hence all debugging messages by using the corresponding preference in Firefox.
- As everyday-users should not be bothered by too many options which are not relevant to them, the debugging mode has to be set directly by accessing the *extensions.webprivacy.debugMode*-preference using “about:config”, Firefoxes’ preference manager.
- P3P policies cannot be fetched directly from the Internet anymore in Firefox 3. This was replaced by using an XML-HTTP-Request.
- Domains with hyphens, numbers etc. can now be checked and IDNs⁶ are also supported.

Basically, Webprivacy builds on three parts of Privacyfox: Fetching the XML-file from the well-known location, checking if it is a reference file or a P3P policy and, more importantly, the parser which translates P3P XML-files into human-readable policies. All of these parts are relevant regarding P3P compliance, that is why they were also enhanced:

- As already described, P3P defines four mechanisms to fetch P3P reference files and Privacyfox only supports the well-known location method. Webprivacy now also has rudimentary support for the other three options build in. Rudimentary because websites which use (X)HTML-links or even HTTP-Headers to provide P3P reference files could rarely be found, so in-depth testing of these features with “real-life” examples was not possible.
- While testing all four mechanisms with as much websites as possible, the author found that a significant number of websites do not correctly implement them: Instead of “w3c/p3p.xml” some other path or filename is used, the P3P HTTP header is either not named correctly (“P3P”) or is invalid, (X-)HTML-link-tags are not named correctly and so on. This is a serious issue for P3P clients as they cannot guess where P3P reference-files can be found. Additionally, they are not allowed to change not well-formed XML-files if they still want to be P3P compliant.
- According to the P3P standard, policies which include the “TEST”-element have to be considered not valid which Webprivacy now does.

So although some of the critical issues regarding P3P compliance have been fixed by Webprivacy, there are still some issues which are not accounted for: Multiple policies in one file, P3P compliant check of the Expiry-element, checking all externally loaded files such as ads or flash-images and cookies and, more importantly: checking if the website

⁶IDN - Internationalized Domain Name

matches the users preferences before it actually loads and can set cookies on the users computer.

Although all those features are clearly important, it has to be mentioned that the aim of this thesis was not to develop a P3P compliant Firefox extension but to provide a basic but user-friendly P3P client which can then be further enhanced by the community.

Furthermore, developing a fully P3P compliant client requires, in the view of the author, at least several man-months to be fully accomplished. In addition to that, it has to be questioned if all requirements of P3P can actually be implemented using a Firefox extension or if a separate plugin would be necessary (see Section 6.1.1 and Section 6.1.9). So although Webprivacy does not fulfill all P3P requirements, it is still a useful, basic P3P client by providing some user-friendly features which will be described next.

6.3.2. Functionality

An important element in designing Webprivacy was to make it as user-friendly as possible. This was accomplished by an easy to use preference system and a toolbar which clearly indicates the current P3P status. With Webprivacy, users can now easily define their own level of privacy by using the “Options” dialog of Firefox. The preferences which can be set are grouped into four thematically clustered areas (Health & Medical, Financial & Purchase, Personally identifiable and Non-Personally identifiable) according to the definition of P3P. The descriptions from each property were taken directly from the P3P standard (cp. [W3Ca, Chapter 6]) and are similar to Privacy Bird, a plugin for Microsoft’s Internet Explorer which was described earlier. In addition to that, users have the option to use pre-defined levels of privacy (low, medium and high) or define a custom level as shown in Figure 6.3.

If a website does not match the users’ privacy preferences and the appropriate status is shown in the toolbar, users of course want to know why the website is violating their privacy preferences. So the original XML-parser from Privacyfox was enhanced by a component which actually checks the P3P policy against the users’ preferences in the first step and, if they do not match, provides a list of reasons why they do not match as shown in Figure 6.4.

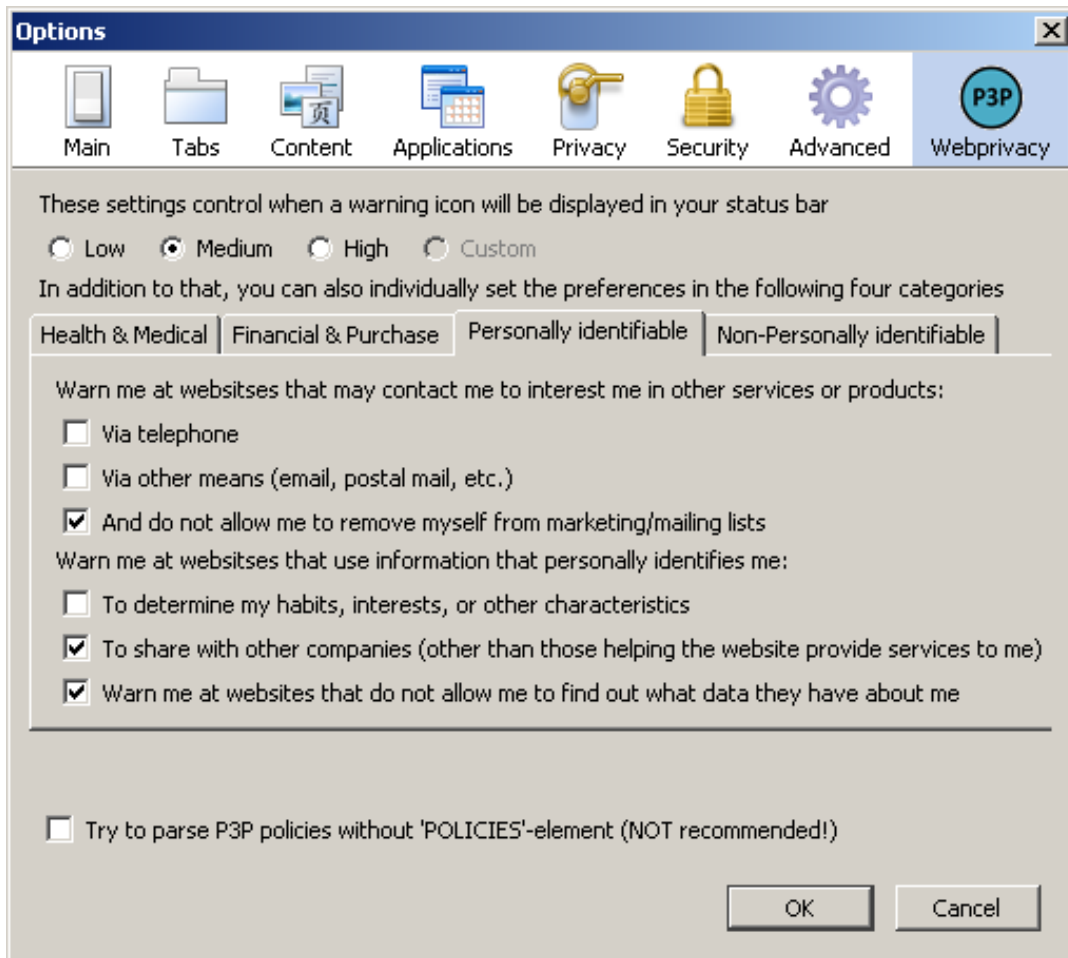


Figure 6.3.: Setting P3P preferences by using the Firefox “Options” dialog

The second important feature of Webprivacy is the status bar on the bottom right corner of the Firefox window. With one icon, the user can immediately check the P3P status which can have one of the following five states:






Icon	State description
	Webprivacy is disabled
	Webprivacy could not find a P3P policy
	The privacy preferences matches the P3P policy of the currently viewed website
	The privacy preferences does not match the P3P policy of the currently viewed website
	The P3P policy which was provided by the website is not well-formed or invalid

Table 6.1.: Five states of Webprivacy

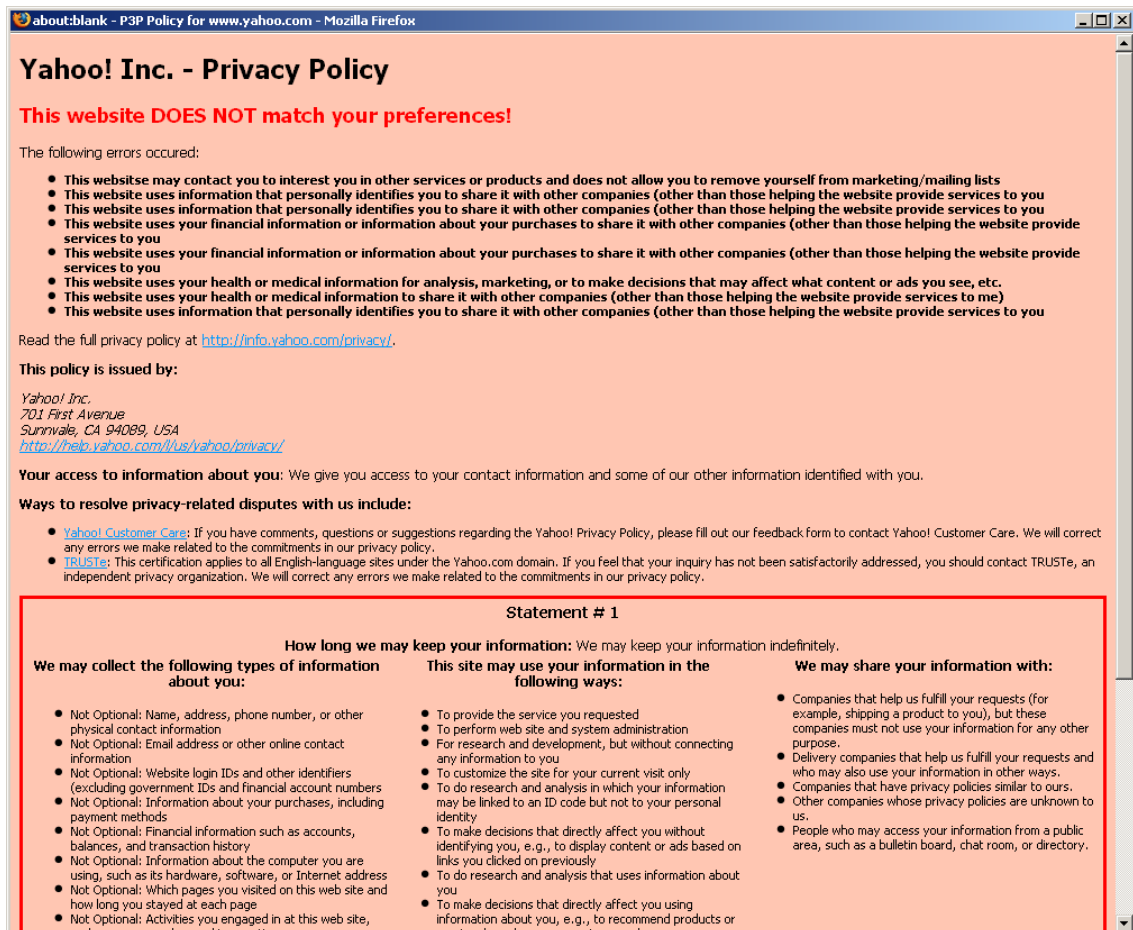


Figure 6.4.: The privacy policy of yahoo.com does not match the users' preferences

Of course, users can also read the human-readable P3P policy of the website as shown in Figure 6.5 by clicking on the status icon. Webprivacy's toolbar also offers a menu which is accessible by right-clicking on the status icon. There, Webprivacy's preferences can be quickly accessed: Webprivacy can be disabled or enabled and Webprivacy's cache can be cleared. Although it is not strictly P3P compliant, Webprivacy caches the results of each check for each domain during a browser session. This was implemented for providing a better user experience: because the P3P policy of a website does not have to be checked more than once per session, the results are available immediately.

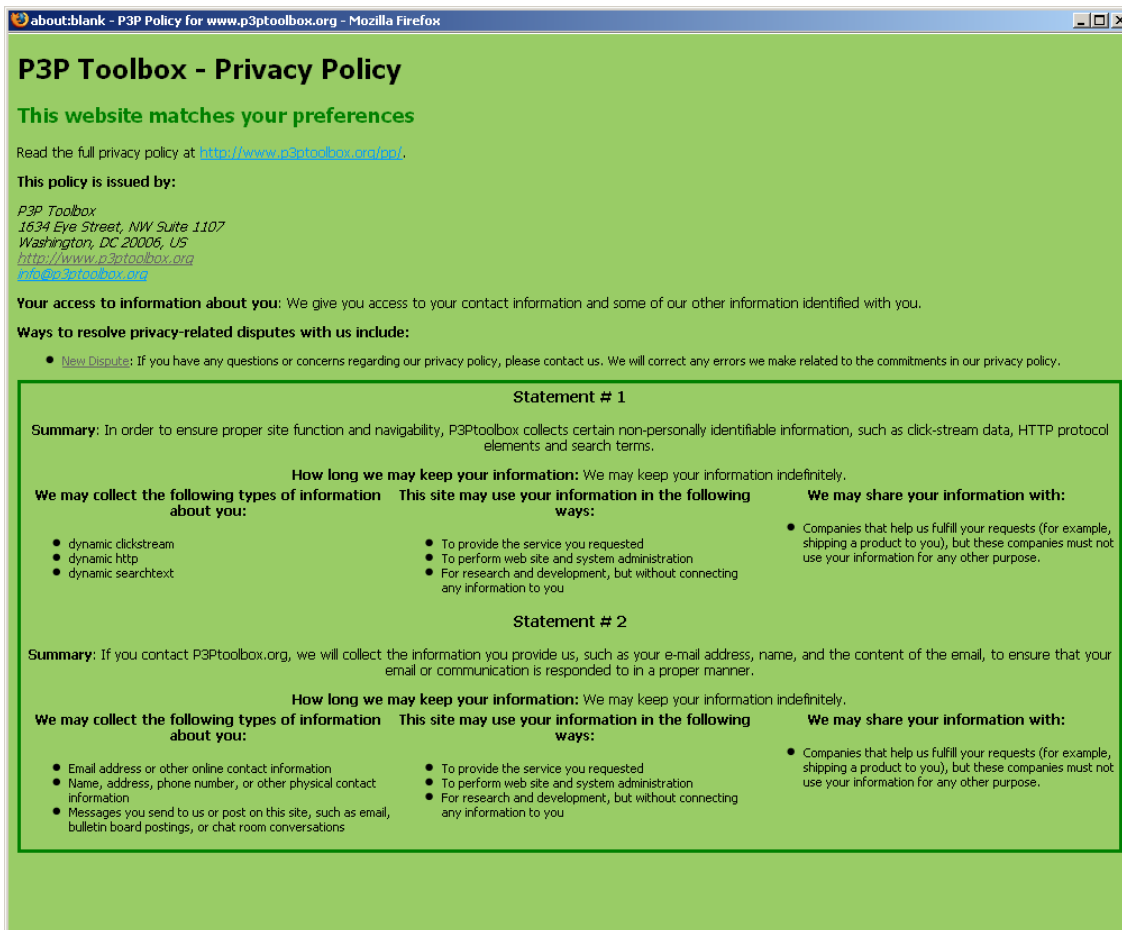


Figure 6.5.: The privacy policy of p3ptools.org does match the users' preferences

6.3.3. Issues

Beside a good user-experience, P3P clients need to be as reliable as possible. However, the P3P standard also has to consider that other organizations or companies have different or more detailed requirements when it comes to expressing their privacy policies. That is why the extension mechanism was introduced to extend the P3P Base Data Schema as already discussed in Chapter 3. However, this raises the question how P3P clients such as Webprivacy should react to data-elements which are not in the Base Data Schema. IBM for example defines its own data schema for data it collects on its website such as the computers' serial number when using the online services for IBM/Lenovo computers.

Another issue was partly addressed in the previous section but should be dealt with in more detail. The problems of not well-formed policies or reference files which link to a non-existing P3P policy should be highlighted on the basis of two examples:

- IBM.com: IBM has a P3P reference file in place at the well-known location, that is "www.ibm.com/w3c/p3p.xml". Besides the issue that this policy reference file

is different from the one available at “ibm.com/w3c/p3p.xml”, the reference file at “www.ibm.com/w3c/p3p.xml” links to non-existing P3P policies at lenovo.com.

- microsoft.com: Microsoft has a valid P3P reference file in place which links to a P3P policy which can be accessed. However, the P3P policy is not well formed. According to P3P v1.1, each policy must have a POLICIES-element which is the root-element of every P3P policy. Unfortunately, this POLICIES-element is missing. The author contacted Microsoft in October 2008 about this issue and received the answer that the message will be forwarded “to the appropriate group for further follow up” (cp. Appendix B) - at the time of writing in February 2009 however, the policy still was not fixed. The P3P policy at amazon.com suffers from the same issue and the author also contacted amazon.com’s customer service and received an answer which could be perceived almost ironically: “Due to the competitive nature of our business, our policy is not to give out information on the inner workings of our company” (cp. Appendix B).

As “forgetting” the POLICIES-element seems to be a common issue, it was decided to implement an option in Webprivacy’s settings to ignore this problem. However, this option is not activated by default and the author does not recommend to use it as the P3P standard explicitly states that “user agents MUST NOT locally modify a P3P policy or policy reference file in order to make it conform to the XML schema” [W3Ca, Chapter 2.4.4]. What is more, it can be questioned if the authors of the P3P policies at these websites created an otherwise valid policy (amazon.com’s policy for example still would be not valid as it lacks almost all basic elements).

The last general issue concerns Firefox itself. Every extension which wants to react to a website being loaded (such as displaying an icon in a toolbar) needs a way to find out when a new website has been opened and finished loading. JavaScript offers so called “event listeners” which fire off a signal which can be traced by extensions. However, opening a website in a browser window fires off up to several dozens of these events as iframes, external resources (such as flash-movies) and even the favicon fires off such an event. That means that an extension which wants to react to websites being loaded (and tabs switched) on the one hand has to find a way not to start too early with its operations and on the other hand only to start its routines once per website. This sounds trivial, but for developing an extension it is not.

6.3.4. Evaluation of the developed add-on

As with the other extensions and plugins in this thesis, Webprivacy will also be evaluated using the criteria outlined in Chapter 4.

Usability and functionality: Webprivacy comes in an .xpi-file and the installation is pretty simple: one only has to drag the xpi-file onto a Firefox browser window and drop it. The software installation manager automatically pops up (cp. Figure 6.6) and asks the user if Webprivacy should be installed. After a restart of Firefox, the Webprivacy toolbar is shown in the bottom right corner of Firefox.

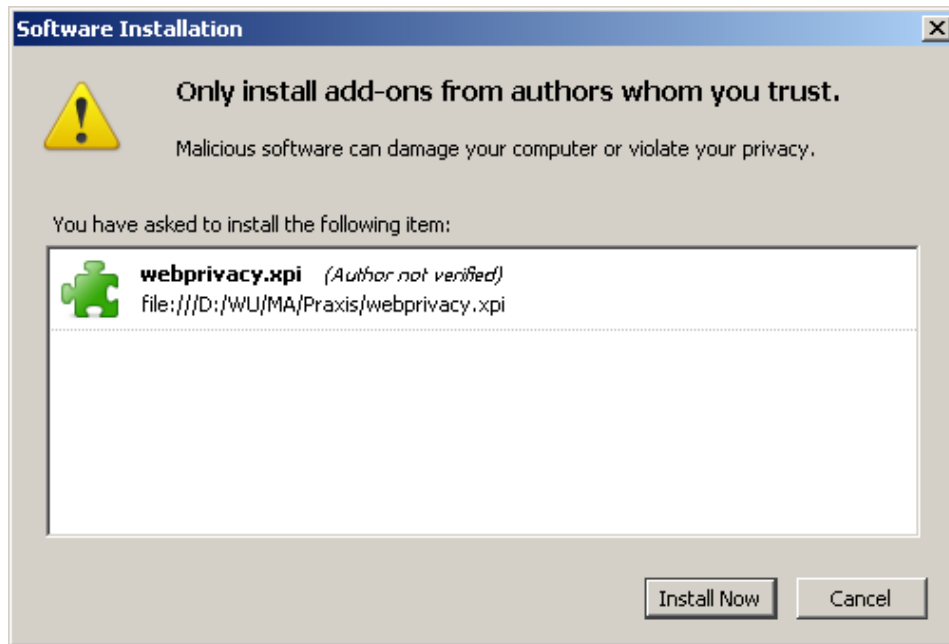


Figure 6.6.: The installation of Webprivacy is done by two clicks

By default, Webprivacy is shipped with the “Medium”-level activated. As Webprivacy’s settings are integrated into the “Options” panel of Firefox, users can easily access them via the “Tools” menu.

While testing Webprivacy, it can be found that Webprivacy very often displays the questionmark-icon, that is Webprivacy was not able to find a P3P policy. This is due to the fact that few websites actually have a P3P policy in place. If a P3P policy can be found and if it is well-formed, Webprivacy often displays the warning-icon to show that this website does not match the users’ preferences. Especially commercial websites often do not match users’ privacy preferences due to their nature of collecting as much data from their users as possible to understand their behaviour and track them. Only privacy related websites such as p3ptoolbox.org often fulfill the “Medium” or “High” setting. This is because most websites collect some kind of information about the users (or they display third-party content which does). For example by displaying advertisements or by using any traffic reporting tool such as Google Analytics. Users could therefore define a custom level and decide which options are most important to them.

When comparing Webprivacy to Privacy Bird, it is found that Webprivacy offers less options, for example does Webprivacy not provide an option to import or export predefined privacy settings.

Reliability and quality in use: Human-readable versions of P3P policies can easily be accessed by clicking on the icon in Webprivacy’s toolbar or using the context-menu. Although Webprivacy provides a quick overview if a policy matches the users’ preferences by using the simple “green = match” and “red = does not match” approach, users still have to read a lot of text to fully understand the privacy policy. Furthermore, Webprivacy is not able to determine a match for non-standard data-elements. This is due to the nature of the extension mechanism: at the moment, Webprivacy can only

access elements in a policy which it recognizes, that is P3P standard elements. This is not the case for elements which have been created by the extension mechanism as the name of the elements and their “meaning” for the policy can be freely defined.

Support of privacy standards: It can also be found that Webprivacy still lacks some of P3P’s features such as multiple policies in one file, support of the expiry-attribute or Compact Policies. However, multiple policies were found not to be widely used and not checking the expiry attribute is not absolutely necessary in the case of Webprivacy as the results of checking the privacy policy are cached for the browser session only.

Prevention of threats: Webprivacy also does not check websites before they are actually loaded or analyzes third-party content on the website (such as images, flash etc.). As was already discussed above, this may be technically not easily possible by using the extension concept. What is more, most P3P enabled websites are currently far away from having their third-party content being P3P enabled - they actually struggle to follow the most basic requirements of defining well-formed XML-files.

Maintainability: As Webprivacy now has a well documented source-code, it should be easy for developers to modify and maintain the software. By providing all necessary build-files with the .xpi-package, developers can quickly start to add new features and build the project.

Overall, Webprivacy is a clear step forward from Privacyfox when it comes to user-friendliness and P3P compliance. However, Webprivacy still is an experimental extension and needs further development to be fully P3P compliant.

7. Conclusion

At the beginning of this thesis the question was raised how Internet users can make sure that their data is only used with their knowledge for the purpose they know and approve of. The answer to this question has been provided throughout this thesis:

Basically, every user leaves traces behind when he or she surfs the Internet. The question is, whether these traces can be used to profile users or even to track them down physically. The answer to this question is “yes”, if users are not concerned enough about their privacy on the Internet, all these things are technically possible. But not all those possibilities are per se “evil”. The Internet revolutionized the way people work, how they shop or pay their taxes. They do this by using online services. These services often need a certain amount of data to function: imaging an e-government service where your identity is not checked. Or a webshop which is not able to deliver the goods purchased because it lacks the information about the home address. The point is that there is a tradeoff between privacy and functionality on the Internet and users have to deal with it and should always be aware of it when they are using privacy sensitive applications.

However, the recent Facebook privacy scandal shows that not all privacy sensitive applications should be used without clearly thinking about what data one wants to share with the application - because dozens of other users or applications may be granted access to this data. Another lesson that one may learn from this recent issue is that the Internet community can be very powerful. Facebook did not inform its users about the change of Terms of Service, however, a team of bloggers found out and within a short period of time, the news spread the Internet, starting from blogs until it reached even the (offline) newspapers. After a huge outcry, Facebook reverted the changes in its Terms of Service due to the public pressure. This is an example where the Internet community successfully controlled itself - but that does not necessarily always has to be the case. And that is why privacy standards and clients which support these standards are necessary.

Ideally, privacy standards make sure that user-data is only used with the knowledge and for the purpose users know and approve of. However, there are four important questions to ask when it comes to privacy standards on the Internet: (1)Are these privacy standards actually widely accepted? (2)If so, are they implemented on a broad basis? (And are they then implemented as designed or is the implementation just a case of window-dressing?) (3)Is there client-software available which actually implements these privacy standards in a user-friendly way? and more importantly (4)Are users actually aware of the issues these standards try to protect them of?

Take P3P for example. The P3P 1.0 recommendation was officially published in April 2002, version 1.1 of the specification in November 2006. Yet, there is no full implementation available although the standard has been available for years. One of

the biggest e-commerce websites on the Internet, amazon.com, is far away from having a valid and meaningful P3P policy in place. And even IBM, which was heavily involved in designing P3P, does not have a P3P policy in place which is all-working. Of course, one could argue that IBM is a huge organization and that there is little chance that the webmasters know the people which were involved in designing P3P. But that exactly is the point: P3P has yet not arrived in the heads of people responsible for users' and customers' privacy. And if people who on a daily basis deal with privacy issues don't know or care about P3P, how should normal Internet users?

If users would know about privacy standards, they could in theory demand an implementation of a certain standard from their software vendors. Unfortunately, this is not the case yet, so software vendors do not have the pressure of implementing privacy standards such as P3P. And although even the W3C acknowledges that there is a lack of support for P3P and therefore decided to suspend the work on it, there is a flicker of hope for P3P: Microsoft Internet Explorer. Maybe in some future release of MSIE, Microsoft will implement a P3P configuration that users can influence. The second flicker of hope for some of the privacy standards described in this thesis is the business community. Maybe not all users are aware of privacy issues, but business certainly are. So maybe the business community develops enough pressure on software vendors to implement useful privacy standards.

Until then, users have to find other solutions to protect their privacy on the Internet. Turning off cookies and not providing any personal data at all is not really an option if one wants to comfortably surf the Internet. That is why tools which at least partly protect users' privacy were introduced and evaluated. Although with hindsight it can be said that the more privacy a tool guarantees, the less comfortable it is to surf the Internet with.

When it comes to P3P, Webprivacy will hopefully do it's bit to the vision of Mozilla Firefox supporting P3P. Webprivacy certainly is neither perfect nor fully P3P compliant. But it is a good starting point for the open source community. That is why the authors' appeal to the community is to download the sourcecode¹ and to further enhance it step-by-step. Maybe then, the value of a privacy standard such as P3P will be discovered and broader support for P3P will be provided.

¹The sourcecode and the ready-to-install xpi-file of Webprivacy are available at <http://svn.semanticlab.net/svn/oss/thesis/webprivacy>

References

- [Ant07] Annie I. Antn, Elisa Bertino, Ninghui Li, and Ting Yu. A roadmap for comprehensive online privacy policy management. *Communications of the ACM*, 50(7):p109 – 116, 2007.
- [Ars04] Fahd Arshad. Privacy fox - a javascript-based p3p agent for mozilla firefox. School of Computer Science, Carnegie Mellon University, December 2004.
- [Awa06] Naveen Farag Awad and M. S. Krishnan. The personalization privacy paradox: An empirical evaluation of information transparency and the willingness to be profiled online for personalization. *MIS Quarterly*, 30(1):p13 – 28, 2006.
- [Bel06] France Belanger and Janine S. Hiller. A framework for e-government: privacy implications. *Business Process Management Journal*, 12(1):p48 – 60, 2006.
- [Boe08] Jorgen Boegh. A new standard for quality requirements. *IEEE Software*, 25(2):p57 – 63, 2008.
- [CNT06] Declan McCullagh. AOL’s disturbing glimpse into users’ lives. http://news.cnet.com/2100-1030_3-6103098.html, 7 August 2006. Accessed 26.07.2008.
- [Cel08] Danielle Celermajer. The State of Free Speech. *Australian Journal of Political Science*, 43(3):p495 – 511, 2008.
- [Coo] Cookiecentral.com. What are cookies? <http://www.cookiecentral.com/cm002.htm>. Accessed on 31.08.2008.
- [Cov08a] Robin Cover. Extensible access control markup language (xacml). <http://xml.coverpages.org/xacml.html>. Accessed 19.08.2008.
- [Cov08b] Robin Cover. Extensible access control markup language (xacml). <http://xml.coverpages.org/xacl.html>. Accessed 19.08.2008.
- [Cra02] Lorrie Faith Cranor. *Web Privacy with P3P*. O’Reilly Media, Sebastopol, CA, USA, 2002.
- [Dow05] Martin R. Dowding. Terrorism, trade, and internet privacy. *Canadian Journal of Communication*, 30(1):p89 – 98, 2005.

- [Dro06] Dennis Drotar, Rachel Greenley, Ahna Hoff, Courtney Johnson, Amy Lewandowski, Melisa Moore, James Spilsbury, Dawn Witherspoon, and Kathy Zebracki. Summary of issues and challenges in the use of new technologies in clinical care and with children and adolescents with chronic illness. *Children's Health Care*, 35(1):p91 – 102, 2006.
- [EPAB09] Andreas Badelt Edin Pezerovic. Cache+. Unpublished seminar-paper written in subject #1902 Seminar aus Informationswirtschaft at WU-Wien (WS08/09) which is available at <http://tmp.badelt.at/Cacheplus.pdf>, January 2009.
- [EU06] European Union. DIRECTIVE 2006/24/EC OF THE EUROPEAN PARLIAMENT AND OF THE COUNCIL. Official Journal of the European Union, March 2006.
- [FSec] Choon Hong. JavaScript Injection Attack. <http://www.f-secure.com/weblog/archives/00001502.html>. Accessed on 26.09.2008.
- [Fel07] Kenneth Feldt. *Programming Firefox: Building Rich Internet Applications with XUL*. O'Reilly Media, 2007.
- [GOA07] Derek Stewart. Elections: Action plans needed to fully address challenges in electronic absentee voting initiatives for military and overseas citizens: Gao-07-774. *GAO Reports*, 6/14/2007:pp1, 2007.
- [Gid06] Julia Gideon, Lorrie Cranor, Serge Egelman, and Alessandro Acquisti. Power strips, prophylactics, and privacy, oh my! In *SOUPS '06: Proceedings of the second symposium on Usable privacy and security*, pages 133–144, New York, NY, USA, 2006. ACM.
- [Gre03] Greg Linden, Brent Smith, and Jeremy York. Amazon.com recommendations. *IEEE Internet Computing*, 7(1):p76–82, 2003.
- [Gro05] Ralph Gross, Alessandro Acquisti, and III H. John Heinz. Information revelation and privacy in online social networks. In *WPES '05: Proceedings of the 2005 ACM workshop on Privacy in the electronic society*, pages 71–80, New York, NY, USA, 2005. ACM.
- [Gua07] Bobbie Johnson. Privacy warning for young users of networking sites. <http://www.guardian.co.uk/technology/2007/nov/23/facebook>, 23 November 2007. Accessed 26.07.2008.
- [Han01] David Hand, Heikki Mannila, and Padhraic Smyth. *Principles of Data Mining*. The MIT Press, Cambridge, MA, 2001.
- [Hil06] Alain Hiltgen, Thorsten Kramp, and Thomas Weigold. Secure internet banking authentication. *IEEE Security and Privacy*, 4(2):21–29, 2006.

- [Hor05] Amir M. Hormozi. Cookies and privacy. *EDPACS: The EDP Audit, Control & Security Newsletter*, 32(9):p1 – 13, 2005.
- [IBM03] International Business Machines Corporation. Enterprise Privacy Authorization Language (EPAL 1.2). <http://www.zurich.ibm.com/security/enterprise-privacy/epal/Specification/index.html>, 14. November 2003. Accessed on 31.03.2009.
- [IEE98] Software Engineering Standards Committee of the IEEE Computer Society. IEEE Std 830-1998: IEEE recommended practice for software requirements specifications. USA, 1998.
- [ISO01a] International Organization for Standardization. ISO/IEC 9126-1:2001: Software engineering – Product quality – Part 1: Quality model. Geneva, Switzerland, 2001.
- [ISO01b] International Organization for Standardization. ISO/IEC 9126-4:2001: Software engineering – Product quality – Part 4: Quality in use metrics. Geneva, Switzerland, 2001.
- [ISO07] International Organization for Standardization. ISO/IEC 25030:2007: Software engineering – Software product Quality Requirements and Evaluation (SQuARE) – Quality requirements. Geneva, Switzerland, 2007.
- [IW01] Jason Levitt. P3P: Protector Of Consumers’ Online Privacy. <http://www.informationweek.com/news/software/development/showArticle.jhtml?articleID=6506308>, 20 August 2001. Accessed on 14.02.2009.
- [Jag07] Tom N. Jagatic, Nathaniel A. Johnson, Markus Jakobsson, and Filippo Menczer. Social phishing. *Communications of the ACM*, 50(10):p94 – 100, 2007.
- [Jon05] Harvey Jones and Jos Hiram Soltren. Facebook: Threats to Privacy. Massachusetts Institute of Technology, 2005.
- [Kah08] Charles M. Kahn and William Roberds. Credit and identity theft. *Journal of Monetary Economics*, 55(2):p251 – 264, 2008.
- [Kha06] Olga Kharif. Big brother is reading your blog. *Business Week Online*, 2/28/2006:p9, 2006.
- [Kit08] Meelis Kitsing. Explaining the e-government success in estonia. In *dg.o ’08: Proceedings of the 2008 international conference on Digital government research*, pages pp429–430. Digital Government Society of North America, 2008.

- [Kob07] Alfred Kobsa. Privacy-enhanced personalization. *Communications of the ACM*, 50(8):p24 – 33, 2007.
- [MDC09a] Mozilla Developer Center (MDC). Extensions. <https://developer.mozilla.org/en/Extensions>. Accessed 15.02.2009.
- [MDC09b] Mozilla Developer Center (MDC). XML User Interface Language (XUL). <https://developer.mozilla.org/en/XUL>. Accessed 15.02.2009.
- [MDC09c] Mozilla Developer Center (MDC). Firefox 3 for developers. https://developer.mozilla.org/en/Firefox_3_for_developers. Accessed 15.02.2009.
- [MDC09d] Mozilla Developer Center (MDC). Building an Extension (Tutorial). https://developer.mozilla.org/en/Building_an_Extension. Accessed 15.02.2009.
- [MDC09e] Mozilla Developer Center (MDC). Install Manifests. https://developer.mozilla.org/en/Install_Manifests. Accessed 15.02.2009.
- [MDC09f] Mozilla Developer Center (MDC). Preferences API. https://developer.mozilla.org/en/Preferences_API. Accessed 15.02.2009.
- [MDC09g] Mozilla Developer Center (MDC). Chrome Registration. https://developer.mozilla.org/en/Chrome_Registration. Accessed 15.02.2009.
- [MDC09h] Mozilla Developer Center (MDC). The Mozilla Developer Center. <https://developer.mozilla.org/En>. Accessed 15.02.2009.
- [MDC09i] Mozilla Developer Center (MDC). Template Guide. https://developer.mozilla.org/en/XUL/Template_Guide. Accessed 15.02.2009.
- [MDC09j] Mozilla Developer Center (MDC). mozStorage API. <https://developer.mozilla.org/en/Storage>. Accessed 15.02.2009.
- [MDC09k] Mozilla Developer Center (MDC). File I/O. https://developer.mozilla.org/index.php?title=File_I%2F%2FO. Accessed 15.02.2009.
- [MS07a] Microsoft Cooperation. The Default Privacy Settings for Internet Explorer 6. <http://support.microsoft.com/kb/293222>, 31 January 2007. Accessed 14.02.2009.
- [MS07b] Microsoft Cooperation. Description of the Platform for Privacy Preferences (P3P) Project. <http://support.microsoft.com/kb/290333>, 31 January 2007. Accessed 14.02.2009.

- [MS07c] Microsoft Cooperation. You Cannot Click "Here" in the Privacy Policy Dialog Box to View a Web Site's Privacy Policy. <http://support.microsoft.com/kb/318810>, 01 February 2007. Accessed 14.02.2009.
- [MSDN] Internet Explorer Developer Center. How to Create a Customized Privacy Import File. <http://msdn.microsoft.com/en-us/library/ms537344.aspx>. Accessed on 14.02.2009.
- [Miy08] Anthony D. Miyazaki. Online privacy and the disclosure of cookie use: Effects on consumer trust and anticipated patronage. *Journal of Public Policy & Marketing*, 27(1):p19 – 33, 2008.
- [Moo05] Trevor Moores. Do consumers understand the role of privacy seals in e-commerce? *Communications of the ACM*, 48(3):p86 – 91, 2005.
- [Moz03] Bugzilla@Mozilla. Remove p3p from the default build. https://bugzilla.mozilla.org/show_bug.cgi?id=225287. Accessed on 14.02.2009.
- [Moz07] Bugzilla@Mozilla. Remove extensions/p3p from the tree. https://bugzilla.mozilla.org/show_bug.cgi?id=366611. Accessed on 14.02.2009.
- [NLC] Susan Haigh and Janette Megarity. Measuring web site usage: Log file analysis. <http://epe.lac-bac.gc.ca/100/202/301/netnotes/netnotes-h/notes57.htm>. Accessed on 31.08.2008.
- [NYT] Brad Stone Brian Stelter. Facebook Withdraws Changes in Data Use. <http://www.nytimes.com/2009/02/19/technology/internet/19facebook.html>. Accessed 23.02.2009.
- [Nam06] Changi Nam, Chanhoo Song, Euehun Lee, and Chan Ik Park. Consumers' privacy concerns and willingness to provide marketing-related personal information online. *Advances in Consumer Research*, 33(1):p212 – 217, 2006.
- [New05] Graeme R. Newman and Megan M. McNally. Identity theft literature review. July 2005.
- [OAS05] Organization for the Advancement of Structured Information Standards. OASIS eXtensible Access Control Markup Language 2.0 (XACML). USA, 2005.
- [OR09] Onion-Info. Onion Routing - Executive Summary. <http://www.onion-router.net/Summary.html>. Accessed 14.02.2009.

- [Oli04] Nadia Olivero and Peter Lunt. Privacy versus willingness to disclose in e-commerce exchanges: The effect of risk awareness on the relative role of trust and control. *Journal of Economic Psychology*, 25(2):243 – 262, 2004.
- [Opp05] Rolf Oppliger. Privacy-enhancing technologies for the world wide web. *Computer Communications*, 28(16):p1791 – 1797, 2005.
- [Pau07] Paul Mooney. Chinese dean demoted over blog. *Chronicle of Higher Education*, 53(30):p44, 2007.
- [Pik2006] George H. Pike. Search engine scrutiny. *Information Today*, 23(3):p19 – 21, 2006.
- [RIET] rietta.com. Extend Firefox: Your Guide to Writing Firefox Extensions. <http://www.rietta.com/firefox/Tutorial/overview.html>. Accessed 15.02.2009.
- [Rin97] David M. Rind, Isaac S. Kohane, Peter Szolovits, Charles Safran, Henry C. Chueh, and G. Octo Barnett. Maintaining the Confidentiality of Medical Records Shared over the Internet and the World Wide Web. *Ann Intern Med*, 127(2):p138–141, 1997.
- [Ros00] Linda Rosencrance. Amazon charging different prices on some dvds. <http://www.computerworld.com/industrytopics/retail/story/0,10801,49569,00.html>, 2000. Accessed on 31.08.2008.
- [Rub08] Ira Rubinstein, Ronald D. Lee, and Paul M. Schwartz. Data mining and internet profiling: Emerging regulatory and technological approaches. *University of Chicago Law Review*, 75:p261–285, 2008.
- [Sai07] Felipe Saint-Jean, Aaron Johnson, Dan Boneh, and Joan Feigenbaum. Private web search. In *WPES '07: Proceedings of the 2007 ACM workshop on Privacy in electronic society*, pages 84–90, New York, NY, USA, 2007. ACM.
- [Sar03] Ravi Sarathy and Christopher J. Robertson. Strategic and ethical considerations in managing digital privacy. *Journal of Business Ethics*, 46(2):p111 – 126, 2003.
- [Sen04] Subhabrata Sen and Jia Wang. Analyzing peer-to-peer traffic across large networks. *IEEE/ACM Transactions on Networking*, 12(2):p219 – 232, 2004.
- [Sid08] Noore Alam Siddiquee. E-Government and Innovations in Service Delivery: The Malaysian Experience. *International Journal of Public Administration*, 31(7):p797 – 815, 2008.
- [Sin07] N. P. Singh. Online frauds in banks with phishing. *Journal of Internet Banking & Commerce*, 12(2):p1 – 27, 2007.

- [Ste07] G. Stermsek, M. Strembeck, and G. Neumann. A user profile derivation approach based on log-file analysis. In *Proceedings of the International Conference on Information and Knowledge Engineering (IKE)*, June 2007.
- [Stu04] William H. Stufflebeam, Annie I. Antón, Qingfeng He, and Neha Jain. Specifying privacy policies with P3P and EPAL: lessons learned. In *WPES '04: Proceedings of the 2004 ACM workshop on Privacy in the electronic society*, pages p35–35, New York, NY, USA, 2004. ACM.
- [Sun05] Anne Anderson. A Comparison of Two Privacy Policy Languages: EPAL and XACML. Technical report, Sun Microsystems Laboratories, 2005.
- [Tan00] Margaret Tan and Thompson S. H. Teo. Factors influencing the adoption of internet banking. *Journal of the AIS*, 1(5):5, March 2000.
- [Ves07] Jari Vesanen. What is personalization? a conceptual framework. *European Journal of Marketing*, 41(5/6):p409–418, 2007.
- [W3Ca] The World Wide Web Consortium (W3C). The Platform for Privacy Preferences 1.1 (P3P1.1) Specification. <http://www.w3.org/TR/2006/NOTE-P3P11-20061113/>. Accessed on 17.07.2008.
- [W3Cb] The World Wide Web Consortium (W3C). RFC 2616: Hypertext Transfer Protocol – HTTP/1.1. <http://www.w3.org/Protocols/rfc2616/rfc2616.html>. Accessed on 16.08.2008.
- [W3Cc] The World Wide Web Consortium (W3C). A P3P Preference Exchange Language 1.0. <http://www.w3.org/TR/2002/WD-P3P-preferences-20020415>. Accessed on 17.08.2008.
- [W3Cd] The World Wide Web Consortium (W3C). The World Wide Web Security FAQ. <http://www.w3.org/Security/Faq/wwwsf2.html>. Accessed on 09.08.2008.
- [W3Ce] The World Wide Web Consortium (W3C). Platform for Privacy Preferences (P3P) Project. <http://www.w3.org/P3P>. Accessed on 23.02.2009.
- [W3S09] w3schools.com. Browser Statistics. http://www.w3schools.com/browsers/browsers_stats.asp. Accessed on 14.02.2009.
- [Wei07] D.J. Weitzner. Google, profiling, and privacy. *IEEE Internet Computing*, 11(6):p95 – 96, 2007.
- [Wie02] Klaus-Peter Wiedmann, Holger Buxel, and Gianfranco Walsh. Customer profiling in e-commerce: Methodological aspects and challenges. *Journal of Database Marketing*, 9(2):p170ff., 2002.

- [Woo06] Jisuk Woo. The right not to be identified: privacy and anonymity in the interactive media environment. *New Media & Society*, 8(6):p949 – 967, 2006.
- [Xu07] Yabo Xu, Ke Wang, Benyu Zhang, and Zheng Chen. Privacy-enhancing personalized web search. In *WWW '07: Proceedings of the 16th international conference on World Wide Web*, pages 591–600, New York, NY, USA, 2007. ACM.
- [ZDN09a] ZDNet. ZDNet Definition for: Proxy Server. <http://dictionary.zdnet.com/definition/proxy+server.html>. Accessed 14.02.2009.
- [ZDN09b] ZDNet. ZDNet Definition for: Anonymous Proxy. <http://dictionary.zdnet.com/index.php?d=anonymous+proxy>. Accessed 14.02.2009.

A. microsoft.com's P3P policy

```
<?xml version="1.0" encoding="UTF-8"?><POLICY xmlns="http://www.w3.org/2000/12/P3Pv1" discuri="http://go.microsoft.com/?linkid=4412894" opturi="https://profile.microsoft.com/RegSysProfileCenter/Infodefault.aspx">
<ENTITY>
  <DATA-GROUP>
    <DATA ref="#business.name">Microsoft Corporation</DATA>
    <DATA ref="#business.contact-info.postal.street">1 Microsoft Way</DATA>
    <DATA ref="#business.contact-info.postal.city">Redmond</DATA>
    <DATA ref="#business.contact-info.postal.stateprov">WA</DATA>
    <DATA ref="#business.contact-info.postal.country">USA</DATA>

    <DATA ref="#business.contact-info.postal.postalcode">98052-6399</DATA>
    <DATA ref="#business.contact-info.online.email">homepage@microsoft.com</DATA>
    <DATA ref="#business.contact-info.online.uri">http://support.microsoft.com/contactus/?ws=mscom</DATA>
  </DATA-GROUP>
</ENTITY>
<ACCESS><all/></ACCESS>
<DISPUTES-GROUP>

<DISPUTES resolution-type="service" service="http://support.microsoft.com/contactus/?ws=mscom" short-description="Microsoft Customer Service">
  <LONG-DESCRIPTION>If for some reason you believe microsoft.com has not adhered to these principles, please notify us by e-mail at homepage@microsoft.com</LONG-DESCRIPTION>
  <REMEDIES>
    <correct/>
  </REMEDIES>
</DISPUTES>
<DISPUTES resolution-type="independent" service="http://www.truste.org/users/watchdog.html" verification="Truste" short-description="TRUSTe Certification">
```

```

<LONG-DESCRIPTION>Microsoft is a premier sponsor of TRUSTe and a
  member of the TRUSTe privacy program, an independent, non-
  profit initiative whose mission is to build users' trust and
  confidence in the Internet by promoting TRUSTe's principles of
  fair information practices.</LONG-DESCRIPTION>
<IMG src="http://www.microsoft.com/library/images/gifs/profilectr/
  Truste.gif" width="91" height="73" alt="TRUSTe: Click to Verify
  " />
<REMEDIES>

  <correct />
</REMEDIES>
</DISPUTES>
</DISPUTES-GROUP>
<STATEMENT>
<PURPOSE>
<individual-decision /><individual-analysis required="opt-out" /><
  pseudo-decision /><pseudo-analysis /><telemarketing required="opt-
  out" /><tailoring /><customization required="opt-out" /><admin /><
  current /><contact required="opt-out" />
</PURPOSE>
<RECIPIENT><same /><ours /></RECIPIENT>
<RETENTION><indefinitely /></RETENTION>
<DATA-GROUP>

  <DATA ref="#user.name.prefix" />
  <DATA ref="#user.name.given" />
  <DATA ref="#user.name.middle" />
  <DATA ref="#user.name.family" />
  <DATA ref="#user.name.suffix" />
  <DATA ref="#user.jobtitle" />
  <DATA ref="#user.home-info.postal" />
  <DATA ref="#user.home-info.telecom.telephone.intcode" />
  <DATA ref="#user.home-info.telecom.telephone.loccode" />

  <DATA ref="#user.home-info.telecom.telephone.number" />
  <DATA ref="#user.home-info.telecom.telephone.ext" />
  <DATA ref="#user.home-info.telecom.fax.intcode" />
  <DATA ref="#user.home-info.telecom.fax.loccode" />
  <DATA ref="#user.home-info.telecom.fax.number" />
  <DATA ref="#user.home-info.telecom.fax.ext" />
  <DATA ref="#user.home-info.telecom.mobile.intcode" />
  <DATA ref="#user.home-info.telecom.mobile.loccode" />
  <DATA ref="#user.home-info.telecom.mobile.number" />

  <DATA ref="#user.home-info.telecom.mobile.ext" />

```

```

<DATA ref="#user.home-info.telecom.pager.intcode" />
<DATA ref="#user.home-info.telecom.pager.loccode" />
<DATA ref="#user.home-info.telecom.pager.number" />
<DATA ref="#user.home-info.telecom.pager.ext" />
<DATA ref="#user.home-info.online" />
<DATA ref="#user.home-info.online.email" />
<DATA ref="#user.home-info.online.uri" />
<DATA ref="#user.business-info.postal" />

<DATA ref="#user.business-info.telecom.telephone.intcode" />
<DATA ref="#user.business-info.telecom.telephone.loccode" />
<DATA ref="#user.business-info.telecom.telephone.number" />
<DATA ref="#user.business-info.telecom.telephone.ext" />
<DATA ref="#user.business-info.telecom.fax.intcode" />
<DATA ref="#user.business-info.telecom.fax.loccode" />
<DATA ref="#user.business-info.telecom.fax.number" />
<DATA ref="#user.business-info.telecom.fax.ext" />
<DATA ref="#user.business-info.telecom.mobile.intcode" />

<DATA ref="#user.business-info.telecom.mobile.loccode" />
<DATA ref="#user.business-info.telecom.mobile.number" />
<DATA ref="#user.business-info.telecom.mobile.ext" />
<DATA ref="#user.business-info.telecom.pager.intcode" />
<DATA ref="#user.business-info.telecom.pager.loccode" />
<DATA ref="#user.business-info.telecom.pager.number" />
<DATA ref="#user.business-info.telecom.pager.ext" />
<DATA ref="#user.business-info.online" />
<DATA ref="#user.business-info.online.email" />

<DATA ref="#user.business-info.online.uri" />
<DATA ref="#user.employer" />
<DATA ref="#user.department" />
<DATA ref="#dynamic.clickstream" />
<DATA ref="#dynamic.http" />
<DATA ref="#dynamic.http.useragent" />
<DATA ref="#dynamic.searchtext" />
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><physical /></CATEGORIES>

</DATA>
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><online /></CATEGORIES>
</DATA>
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><uniqueid /></CATEGORIES>
</DATA>

```

```

<DATA ref="#dynamic.miscdata">
  <CATEGORIES><purchase /></CATEGORIES>

</DATA>
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><computer /></CATEGORIES>
</DATA>
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><navigation /></CATEGORIES>
</DATA>
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><interactive /></CATEGORIES>

</DATA>
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><content /></CATEGORIES>
</DATA>
<DATA ref="#dynamic.miscdata">
  <CATEGORIES><preference /></CATEGORIES>
</DATA>
<DATA ref="#thirdparty.name.prefix" />
<DATA ref="#thirdparty.name.given" />

<DATA ref="#thirdparty.name.middle" />
<DATA ref="#thirdparty.name.family" />
<DATA ref="#thirdparty.name.suffix" />
<DATA ref="#thirdparty.home-info.postal" />
<DATA ref="#thirdparty.home-info.postal.name" />
<DATA ref="#thirdparty.home-info.postal.street" />
<DATA ref="#thirdparty.home-info.postal.city" />
<DATA ref="#thirdparty.home-info.postal.stateprov" />
<DATA ref="#thirdparty.home-info.postal.postalcode" />

<DATA ref="#thirdparty.home-info.postal.country" />
<DATA ref="#thirdparty.home-info.postal.organization" />
<DATA ref="#thirdparty.home-info.telecom.telephone" />
<DATA ref="#thirdparty.home-info.telecom.telephone.intcode" />
<DATA ref="#thirdparty.home-info.telecom.telephone.loccode" />
<DATA ref="#thirdparty.home-info.telecom.telephone.number" />
<DATA ref="#thirdparty.home-info.telecom.telephone.ext" />
<DATA ref="#thirdparty.home-info.telecom.telephone.comment" />
<DATA ref="#thirdparty.home-info.telecom.fax.intcode" />

<DATA ref="#thirdparty.home-info.telecom.fax.loccode" />
<DATA ref="#thirdparty.home-info.telecom.fax.number" />
<DATA ref="#thirdparty.home-info.telecom.fax.ext" />

```

```

<DATA ref="#thirdparty.home-info.telecom.mobile.intcode" />
<DATA ref="#thirdparty.home-info.telecom.mobile.loccode" />
<DATA ref="#thirdparty.home-info.telecom.mobile.number" />
<DATA ref="#thirdparty.home-info.telecom.mobile.ext" />
<DATA ref="#thirdparty.home-info.online" />
<DATA ref="#thirdparty.home-info.online.email" />

<DATA ref="#thirdparty.home-info.online.uri" />
<DATA ref="#thirdparty.employer" />
<DATA ref="#thirdparty.department" />
<DATA ref="#business.name" />
<DATA ref="#business.department" />
<DATA ref="#business.contact-info.postal" />
<DATA ref="#business.contact-info.postal.street" />
<DATA ref="#business.contact-info.postal.city" />
<DATA ref="#business.contact-info.postal.stateprov" />

<DATA ref="#business.contact-info.postal.postalcode" />
<DATA ref="#business.contact-info.postal.country" />
<DATA ref="#business.contact-info.postal.organization" />
<DATA ref="#business.contact-info.telecom.telephone.intcode" />
<DATA ref="#business.contact-info.telecom.telephone.loccode" />
<DATA ref="#business.contact-info.telecom.telephone.number" />
<DATA ref="#business.contact-info.telecom.telephone.ext" />
<DATA ref="#business.contact-info.telecom.fax.intcode" />
<DATA ref="#business.contact-info.telecom.fax.loccode" />

<DATA ref="#business.contact-info.telecom.fax.number" />
<DATA ref="#business.contact-info.telecom.fax.ext" />
<DATA ref="#business.contact-info.telecom.mobile.intcode" />
<DATA ref="#business.contact-info.telecom.mobile.loccode" />
<DATA ref="#business.contact-info.telecom.mobile.number" />
<DATA ref="#business.contact-info.telecom.mobile.ext" />
<DATA ref="#business.contact-info.telecom.pager.intcode" />
<DATA ref="#business.contact-info.telecom.pager.loccode" />
<DATA ref="#business.contact-info.telecom.pager.number" />

<DATA ref="#business.contact-info.telecom.pager.ext" />
<DATA ref="#business.contact-info.online" />
<DATA ref="#business.contact-info.online.email" />
<DATA ref="#business.contact-info.online.uri" />
</DATA-GROUP>
</STATEMENT>
</POLICY>

```

Listing A.1: microsoft.com's P3P policy (Source: microsoft.com)

B. E-Mail correspondence

Subject: Your Amazon.com Inquiry
Date: Mon, 6 Oct 2008 04:46:49 -0700 (PDT)
From: Amazon.com Customer Service <cust.service03@amazon.com>
To: xxxxx@badelt.at <xxxxx@badelt.at>

Thanks for writing to us at Amazon.com.

Due to the competitive nature of our business, our policy is not to give out information on the inner workings of our company.

You are welcome, however, to review the materials we have prepared for the public. Our press materials and investor information are available online at:

<http://www.amazon.com/ir>

These materials may be able to provide you with the information you need. Otherwise, we hope you'll understand our position. Thank you again for your interest in Amazon.com.

Please let us know if this e-mail resolved your question:

If yes, click here:

<http://www.amazon.com/rsvp-y?c=rbwhfbxd3479770603>

If not, click here:

<http://www.amazon.com/rsvp-n?c=rbwhfbxd3479770603>

Please note: this e-mail was sent from an address that cannot accept incoming e-mail.

To contact us about an unrelated issue, please visit the Help section of our web site.

Best regards,

Subramaniam V.
Amazon.com Customer Service

Check your order and more: <http://www.amazon.com/your-account>

----- Original message: -----

10/05/08 13:21:07

Your Name: Andreas Badelt

Comments: Hello ,

I am currently writing my masters thesis about P3P and I am checking P3P policies of big websites. I found that your P3P policy seems to be not well formed. According to W3C's P3P v1.1 (<http://www.w3.org/TR/P3P11/#POLICIES>) each policy must have a POLICIES element which your policy at <http://www.amazon.com/w3c/p3p-full.xml> is lacking. If you are still testing your policy, there is a TEST-element to indicate that your policy should not be considered valid.

Best Regards from Vienna,
Andreas

Listing B.1: E-mail correspondence with amazon.com's customer service

----- Original Message -----

Subject: RE: SRX1080802937ID - Not wellformed P3P policy at microsoft.com

Date: Mon, 6 Oct 2008 15:02:32 -0700

From: Microsoft Online Customer Service

<CNTUS.PRC.S.NA.00.EN.TRA.BGL.CS.T01.CUS.00.WB@css.one.microsoft.com>

To: <xxxxxx@badelt.at>

Hello Andreas ,

Thank you for contacting Microsoft Customer Service .

Thank you for taking time to provide feedback about Microsoft P3P policies . I am forwarding your message to the appropriate group for further follow up . Your feedback is very important to us . Microsoft is committed to customer satisfaction , and it is only with the help of our valued customers that we can achieve this goal . We strive to provide

the

Thank you ,

Renukesh

Microsoft Customer Service Representative

If you have any feedback about your Online Customer Service experience , please send them to my manager, Justus Joy, at <http://go.microsoft.com/?linkid=6998852> Please do not forget to indicate the name of my manager in the subject field.

— Original Message —

From : xxxxxxxx@badelt.at
Sent : Sunday, October 05, 2008 8:22:47 PM UTC
To : CNTUS.PRIV.WW.00.EN.000.000.CS.CMR.CUS.00.WB@css.one.microsoft.com
Subject : Not wellformed P3P policy at microsoft.com

QUESTIONS OR COMMENTS

Message: Hello ,

I am currently writing my masters thesis about P3P and I am checking P3P policies of big websites. I found that your P3P policy seems to be not well formed. According to W3C's P3P v1.1 (<http://www.w3.org/TR/P3P11/#POLICIES>) each policy must have a POLICIES element which your policy at <http://www.microsoft.com/w3c/p3policy.xml> is lacking. If you are still testing your policy , there is a TEST-element to indicate that your policy should not be considered valid.

Best Regards from Vienna ,
Andreas

Listing B.2: E-mail correspondence with Microsofts's customer service