Browser eavesdropping

how can we prevent our browsers from
revealing our private information

Randi Gjerde

Master's Thesis
Master of Science in Information Security
30 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2005

The MSc programme in Information Security is run in cooperation with the Royal Institute of Technology (KTH) in Stockholm.

Institutt for
informatikk og medieteknikk
Høgskolen i Gjøvik
Postboks 191
2802 Gjøvik

Department of Computer Science
and Media Technology
Gjøvik University College
Box 191
N-2802 Gjøvik
Norway

Abstract

Our browsers can be used as tools to find information about us, map our web browsing habits, or even worse: find our user-names and passwords. This thesis looks at what kind of information it is possible to find about a user just by using hidden programs and scripts in the source code of a web page. We have also tested the possibility of discovering whether a web site is storing information about visitors. Software which claims to protect the identity of Internet users has been tested, to ascertain to what extent Internet users can really be protected from this kind of exploits.

As it was not possible to detect if web sites were storing information about visitors, Internet users have to rely on protection software if they wish to remain anonymous on the Internet. The software tested protected the most important information from being revealed, and will help the users remain hidden from profiling and data mining. However, in most online transactions it is required that the customers give their name and address. The software would in these cases be of no use.

Keywords: web use, trace cookies, browser eavesdropping, information security, Internet anonymising, privacy protection, tracks, hide identities, borderless community.

Sammendrag

Våre nettlesere kan benyttes som verktøy for å finne informasjon om oss, kartlegge våre surfevaner, eller enda verre: finne brukernavn og passord. Denne masteroppgaven ser på hvilken informasjon det er mulig å finne om en bruker bare ved å benytte skjulte programmer og scripts i kildekoden til en webside. Vi har også testet muligheten for å oppdage om en webside lagrer informasjon om besøkende. Programvare som hevder å beskytte identiteten til Internett-brukere har blitt testet for å fastslå i hvilken grad Internett-brukere virkelig kan beskyttes mot denne typen utnyttelse.

Siden det ikke var mulig å detektere om en webside lagrer informasjon om besøkende, må Internett-brukere stole på beskyttelsesprogramvare om de ønsker å forbli anonyme på Internett. Programvaren som ble testet beskyttet den viktigste informasjonen fra å bli avslørt, og vil hjelpe brukere å forbli beskyttet mot profilering og datamining. Likevel er det i de fleste transaksjoner på Internett nødvendig for kunder å oppgi navn og adresse. I slike tilfeller er programvaren til liten eller ingen nytte.

Preface

This Master's thesis has been written the final semester of the Master studies of Gjøvik University College and Kungliga Tekniska Högskolan.

As my focus has been with 'ordinary' Internet users, it was important to choose a subject relevant not only to the information security business, but also for other Internet users. The protection of personal data has been more focused on the last year, as, among other things, 'phishing' scams has become a part of our everyday life. This thesis tries to find solutions to how to protect Internet users from revealing unnecessary information, as well as create more awareness of the amount of information that could be exploited when we browse the Internet.

Some international regulations and guidelines are mentioned in chapter 2. To ease the reading they are only mentioned in an abbreviated form. The full names are available in Appendix 1 – Terminology.

I would like to take the opportunity to thank my thesis supervisor at Gjøvik University College, Lasse Øverlier, for evaluations and input on language, facts and problems. I would also like to mention my other lecturers, co-students and friends who have given me assistance when I encountered problems, as well as useful input on facts and language, encouragement, and moral support during these months: Slobodan Petrovic, Hanno Langweg, Øyvind Kolås, Rune L. Skår, Ketil Gjerde, Henning Gravnås, Christian Bunes, Jan Vidar Simonsen, Britt Karin Rotmo and Monica Strand.

Gjøvik, June 30 2005

Randi Gjerde

Table of Contents

List of Figures

1 Introduction

Every day, as we open our web browsers to perform our tasks, we leave behind an endless amount of tracks, visible to anyone who wants to know who we are, where we are, what we do, and when we do it. Regardless of what most of us may think, anonymity on the Internet is usually not something easily achieved. Using easy methods implemented on any web site, information about a visitor's computer, geographic location and the web pages visited lately can be revealed. Even the information in the computer's clipboard is accessible.

This information can be used and misused. Marketing agencies might track our web habits to see which pages are worth advertising on, or even adapt the web pages to what seems to be the general area of interest. News web sites could be adapted to view local news on top, based on your geographical location. These examples would normally be seen as positive adaptations. But what if someone is listening to see whether you copy a password, or maybe an account number, into your computer's clipboard? How easy would it be to make such an application and insert it hidden onto a website? Is this something which could be done by any programming novice with malicious intent (also known as 'script kiddies')?

Finding information on a specific subject could be quite easy using the Internet, e.g. tutorials on how to find information about visitors to a web page. The question remains whether our browsers could be used as tools by those who want to find out anything about us, and whether it is possible to protect our privacy or not. Through a laboratory test, it has been demonstrated what kind of information it is possible to store about visitors to a web page. A software test has also been performed, to find out how much of the information can be protected.

The thesis will prove that a significant amount of information can be extracted from a visitor's browser, just by using the source code of a web site. It will also demonstrate how it is possible to protect oneself, by using commercially available software.

The following research questions were defined and have been answered in this thesis:

- What kind of information is it possible to expose and store about Internet users without their knowledge just by exploiting browser vulnerabilities?
- How easy is it to locate and store this information?
- Is it possible to detect whether a web site collects unnecessary information about visitors?
- What can be done to protect against this information getting exposed and stored?
- How do current laws regulate the collection and use of information about Internet users without their knowledge?

The following will give an introduction to the structure of this thesis. Chapter 2 explains the theory and background for the thesis, and looks at what research has already been undertaken in relevant areas. It also looks into the legal aspects of the collection and use of information about Internet users, especially with regards to the difference between European and US regulations. Chapter 3 explains the approach with which the results were obtained in the laboratory test, as well as the work necessary to save information about the visitors to the test web page. Chapter 4 explains the approach of the information detection and privacy software tests, as well as the results. In Chapter 5 we discuss the results of the experiment and the tests, and give recommendations on how to take the results into consideration. In Chapter 6 the thesis concludes, and suggests further work on relevant subjects. Appendix A offers explanations to the terminology and abbreviations used in the thesis. In Appendix B the invitation to the participants of the laboratory test is shown, as well as the test web page itself and the results from the test displayed in table format.

2 Theory and background

The main focus of this thesis is to raise the level of privacy for average Internet users by doing research on what kind of information is revealed by the web browser, and testing software which is supposed to protect against this information being revealed. It is demonstrated how much information it is possible to extract about a person visiting a web site just by using hidden programs and scripts in the source code of a web page. It is also important to demonstrate whether it is possible to detect if the browser is revealing information, e.g. by sending it to a database or storing it to cookies. The thesis has been developed according to methods described by Creswell [22], Booth et al. [10] and Salkind [70] .

2.1 Motivation

There has been nothing in the history of humankind like the Internet [47]. Nothing else can give access to such a vast amount of information; nothing else enables the user to communicate in so many ways with so many other people whether they be down the street or on the other side of the globe. Nothing else gives such power over the way we shop, the way we live, the way we work, and the way we have fun. And nothing else can allow our privacy to be invaded or security threatened so easily [47].

Exchanging personal information in return for services that are more targeted, convenient or useful, has become a part of everyday life. All sorts of services now require us to give some information about ourselves to participate, whether it's voting, driving, travelling, shopping or subscribing to a magazine [56]. The higher the levels of service customers demand, the more information they may have to provide to get the required service [69]. People often have very little control over what their personal information is used for afterwards [56]. In addition to the customer information that is voluntarily provided by the customers themselves, businesses can also collect information on customer online behaviour using cookies and click-stream analysis [69]. The necessity for protection of user privacy on the Internet could be hard to understand for the 'normal Internet user', the opinion is often that there is nothing to hide [25]. However, the Internet continues to raise awareness of privacy concerns because of the massive amount of personal information that can be collected, shared and disclosed [19].

2.1.1 Security and privacy

According to Clayton and Stewart [19], privacy and security are not the same thing. Privacy is concerned with an individual's ability to control how an organization gathers, uses and discloses personally identifiable information. Security refers to how an organisation protects the data during and after collection [19].

It is important to protect privacy, even for the average Internet user. Anonymity is one important form of privacy protection that is often useful [46]. As for now, the privacy is usually not protected without the user taking some kind of measures about protecting oneself. Even in these cases one is normally not guaranteed to be completely secure, as the ‘anonymising’ programs are mainly recommended by the very company which market them. It is also possible to hide tracking programs, which sole purpose is to send information about surfing habits to a database, inside other programs.

There is an increasing need to assess the possibilities we have to hide our information from getting revealed. There are two kinds of information given up by visitors to a web page: protocol information exchanged between the web browser and the web server, and personal information given up by the user [25]. If individuals have not actively disclosed information about themselves, they believe that no one knows who they are or what they are doing [9]. Anyone using the Internet should be aware of what kind of information could be and often is collected about them. We will in this thesis concentrate on the first kind of information mentioned here: protocol information exchanged between the web browser and the web server.

2.1.2 Work done

The thesis demonstrates what kind of information can be extracted about a visitor to a web page, and how easy it is to do the programming necessary to extract the information. It also assesses the possibilities we have to hide our information, for example by installing software which protects the privacy of the user. There are continually newspaper articles claiming to know how to avoid leaving tracks, but the articles do not always have trustworthy sources for their claims. There seems to be few tests on how good this commercially available software really is. It would be important to know if it is of any use to install the software, and if it really protects the privacy of the user. This thesis assesses some of the ‘anonymising’ software available, to decide how and to what extent they protect against the information being revealed.

An important part of the process of hiding information is trying to detect whether a web page is collecting information about the visitor. There has been research done in this thesis on how one could try to detect this, and if software exists for this purpose. The thesis also looks further into what it will take to detect eavesdropping like this, as there seems to be a lack of information available on how to detect when information is being collected.

2.2 Previous research

In recent years the usage of the Internet (the *World Wide Web*) has increased tremendously. It has developed from a service focused on academic areas offering scientific content, into a medium for providers of information of different kinds and often doubtful seriousness [27]. Based on the fact that a user on the Internet gives away a lot of information while navigating it, it is possible to build a personal profile of this user. The accumulation and connection of this information contradicts the usual idea of data security, violates personal rights, and offers many illegal possibilities like insertion into unwanted address lists, which often results in undesired advertisements or spying out personal tendencies [27]. Because of this situation, some services offer users to access web pages unrecognised or without the risk of being backtracked. This could for example be by the use of a proxy [26] or according to Miller [58, p. 410], Levine et al. [57, p. 146] and Schaeffer [72, p. 124], a service like `www.anonymizer.com`.

According to Rust et al. [69], science fiction writer Isaac Asimov and political novelist Ayn Rand take opposite sides in the view of the eventual outcome with respect to privacy. According to Asimov, ‘the advance of civilisation is nothing but an exercise in the limiting of privacy’. However, Rand says that ‘civilisation is the progress toward a society of privacy’. Rust et al. [69] argue that if privacy is left to market forces, the future will be a mix of the Asimov and Rand points of view. That is, privacy will continue to decline, but it will not go away because the emerging privacy industry will persist indefinitely, although it is likely to shrink over time as the maintenance of privacy becomes more expensive [69].

2.2.1 Privacy policies

According to Antón and Earp [6], Internet users are more inclined to trust a web site if it posts a privacy policy. A privacy policy is a comprehensive description of a web site’s information practices, located in an easily accessible place on the site [6]. Consumers often have only the stated web site policies as a guide to how their information is used, and thus on which to base their browsing and transaction decisions.

The Electronic Privacy Information Center (EPIC) reviewed 100 of the most frequently visited web sites on the Internet in 1997 [35], 1998 [36] and 1999 [37], to check on how the web sites handled privacy issues. The focus was on establishing if the sites collected personal information, had published privacy policies, made use of cookies, and allowed people to visit without disclosing their actual identity. They found that in 1997, few web sites had created explicit privacy policies, and none of the samples met basic standards for privacy [35]. In 1999 more sites were posting privacy policies, as the rise of new associations to promote the development of privacy policies and encourage industry awareness of privacy issues was seen [37]. However, the privacy policies found were often incomplete, especially with regards to the intended use of the information collected, and who would have access to this information. Simultaneously marketers were using new and more sophisticated techniques to track consumers on the Internet. In the online world, every consumer inquiry about a product and every viewing may quickly become incorporated into a detailed profile that will remain hidden from the consumer [37].

The Platform for Privacy Preferences (P3P) is a standard computer-readable format for online privacy policies developed by the World Wide Web Consortium (W3C) [21]. P3P-encoded privacy policies can be fetched automatically by P3P-enabled web browsers and other P3P software. P3P is a standardized set of multiple-choice questions, covering all the major aspects of a Web site’s privacy policies. Taken together, they present a clear snapshot of how a site handles personal information about its users [77]. According to Cranor [21], as of July 2002 basic P3P functionality had been built into the Microsoft Internet Explorer 6 and Netscape Navigator 7 web browsers. However the P3P implementations in these browsers are limited to automated processing of cookies and display of summary privacy policies when requested by a user.

P3P adoption by web sites has proceeded at an encouraging pace since P3P became an official W3C Recommendation in April 2002. However, as P3P adoption is entirely voluntary, many sites do not view P3P as a high priority [21]. Because of the increased transparency that comes as a result of companies using P3P, some improvements in web site privacy policies are likely. According to Cranor [21], P3P may indirectly help to improve web site privacy policies in jurisdictions where there are few legal requirements for privacy policies. Anyhow, P3P allows companies to communicate their privacy policies, but it offers no guarantee that companies will actually follow their policies, it relies on site owners' participation and honesty [6]. According to Antón and Earp [6], a report by EPIC asserts that P3P fails to comply with baseline standards for privacy protection, and that it is a complex and confusing protocol that will hinder Internet users in protecting their privacy. However, by P3P enabling the privacy policy, a company displays their willingness to at least make an effort.

2.2.2 Trust

According to DeVault et al. [28], consumers are looking for evidence of a company's efforts to build trust with them. The ease of data collection and ongoing revelations about how and where people's privacy expectations are not being met, have made privacy the most important trust issue for online businesses. In their online activities, consumers inquire about company privacy policies that describe the use of consumer information. Over the past years, businesses have increased their efforts to build trust with consumers by investing in earning consumers' confidence with respect to privacy. According to DeVault et al. [28], even after several years of business focus on addressing privacy concerns, a majority of consumers continue to have significant concerns that businesses are not keeping the promises they make in their privacy policies.

Browser-based technologies, e.g. P3P, that assist consumers in evaluating privacy policies, still fail to address the root of the privacy trust issue. Consumers do not believe companies are doing what they say they are doing. These technologies merely automate a manual process that consumers distrust.

The European Commission performs surveys regularly on consumer behaviour and rights. They concluded as early as in 1996 that consumers do not like that information is being collected on the Internet [31]. The survey showed that two thirds of all Europeans were worried about the tracks they might leave by using information networks. In 2003 the European Commission surveyed the level of trust consumers had in the use of personal data held by organisations and services such as banks, police, doctors, and so on [32]. This report reveals that the share of people who are worried about the protection of privacy is about the same as in 1996.

A survey performed in 2004 on E-commerce states that Internet commerce had then been used by 16% of EU citizens [33]. One major interesting element coming out of this research is that the most important limiting factor affecting the e-commerce market is, in fact, neither confidence nor issues such as security of payment, language, etc. but the fact that 57% of EU citizens are not connected to the Internet and, therefore, do not have the means to undertake e-commerce [33].

According to a survey performed by TNS Norsk Gallup [40] in 2000, only around 63% of Norwegians had Internet access, and out of these, 38% used the Internet for banking, and around 19% used the Internet for shopping. Those using the Internet for shopping cite *simplicity of use* as the most important reason for shopping online. The most important reasons for not shopping online seem to be *no need* and *lack of security*. This is supported in the 2004 report by the European Commission, where lack of *trust* and *interest* seem to be the two major reasons for not shopping on the Internet, not counting those with no access to the medium.

According to the Electronic Frontier Foundation (EFF), surveys show that over 66% of people who have not used the Internet would be more likely to start using it if the privacy of their personal information and communications would be protected [56]. Clayton and Stewart [19] claim that if privacy is not managed or is ignored, companies may risk loss of consumer and employee goodwill, money wasted on avoidable privacy litigation, bad publicity and decreased stock value. On average, in 2003, 60% of all EU citizens were concerned about the protection of privacy to a greater or lesser degree [32]. The same survey reveals that nine out of ten EU citizens want to be informed why organisations are gathering their personal data, and whether the data are being shared with other organisations.

2.2.3 Choice

One of the issues in the privacy debate has been the manner in which a consumer can make choices about the use of their personal information [55]. The debate is regarding the opt-out or opt-in options. According to Abrams [2], in the USA consumers designate choice through an opt-out mechanism, while in Europe individuals are given the choice to opt in.

Opt-out means that a business is free to use consumer information for marketing and selling unless the consumer objects. *Opt-in* means that a business is prohibited from using information for selling and marketing without the consumer's explicit approval [2]. Because consumers may not be aware of their options under an opt-out regime, most privacy bills that have called for opt-out have also required clear and conspicuous notice of consumer choices about data sharing. According to Lawler and Cooper [55], most consumer advocates prefer opt-in regimes, as opt-in requires businesses to both inform and ask consumers to make an active choice in agreeing to share their personal information.

2.3 Technologies affecting privacy

The communication between web browser and web server works in a bilateral way. The browser initiates a request consisting of two parts, the header and the body. The HTTP header contains meta information and data fields specifying the address to be contacted. According to Demuth and Rieke [26] the header may contain the type of browser the client uses, the operating system, the IP address of the computer, the geographic location of the Internet access, whether the web server has been contacted earlier, and the address of the previous web page visited. The actual content of the web page is transported in the body, e.g. graphics and parameters from forms. After receiving a request the web server reacts with a response which is constructed similarly [27]. The request consists of normal ASCII text, and can therefore easily be read by humans [25].

While browsing web pages, a surfer is implicitly presenting the addresses of the visited pages to various instances. Common web browsers are logging, or 'caching', used web addresses in a so called 'URL history', offered to the user to simplify the access to the user's recently visited web pages. A URL can also carry additional information. Password, context, or personal information could be coded in the URL rather than using other methods to maintain state (see Chapter 2.3.1 for more information on maintaining state).

Furthermore it is possible that copies of requested pages are available at all instances between the web server and the client. Network nodes cache web pages to be able to deliver often requested pages in a shorter time. [27]

One HTTP header is of particular interest when dealing with security, for a couple of reasons [51]. The header is named *Referer*. A Referer header is sent by most browsers on most requests. The header contains the URL of the document from which the request originated. It is sent to contact the sites which the originator document links to. One of the problems with the Referer header, from a security point of view, is that it leaks information to remote sites. Any part of the URL, including parameters, will be visible to the third-party web server and any proxies that handle the request. Another problem with the Referer header is that it originates on the client. This makes it possible to modify the Referer header to make it look like it comes from another site [51].

The Referer header helps website owners evaluate how visitors arrived, for example via links or search engines, as this is coded into the header [59]. The privacy implications of the Referer data are clear. Actually, the accepted standard for the HTTP protocol makes special mention of such privacy issues [11]: 'Because the source of a link may be private information or may reveal an otherwise private information source, it is strongly recommended that the user be able to select whether or not the Referer field is sent'. Yet, according to Broder [11], only the Opera browser implements the recommendation to allow users to disable Referer transmission.

2.3.1 Cookies

HTTP is a 'stateless' protocol. A client sends a request, the server responds, and then both forget that they have talked to each other [51]. We would, however, often like to have 'state' between requests. Cookies were introduced as an extension to HTTP to create a possibility to maintain that state. With cookies, the web server asks the client to remember a small piece of information. This information is passed back by the client on each subsequent request to the same server for as long as the cookie lives, to say that there has been previous communication. A session cookie is erased once the browser is closed. A cookie that expires after some time can be accessed any time until it expires, except in cases where the browser overrides server wishes and expires the cookies as soon as the browser is closed. The client has no idea what the information means, it just knows that the server needs this information, and faithfully passes it back [51].

Cookies tend to be misperceived and seen as dangerous, but cookies are harmless by nature. In no way, shape or form can a cookie scan a user's hard drive, or collect a credit card number [4]. A cookie is simply a value that a server requests to be stored on the client. Most browser implementations save cookies as small plain-text files. The type of data stored in the cookie depends on the supplier of the web site. According to Schaeffer [72], it very often simply records when you visited the particular web page. To make your future online orders easier, data from form fields, for instance your address or your credit card number might also be stored. The next time an order is placed, the data is automatically entered into the form.

Many web sites use cookies to help gain access to advanced features of that site. This includes personalised features from the news and stock preferences, to the horoscope. Shopping carts can use cookie features to keep track of what products you have placed in the cart or identify you such that your selections may be retrieved later [42]. If the acceptance of cookies is turned off, many sorts of functionality will be denied.

Amazon (`http://www.amazon.com`) offers customers a way to personalise their shopping experience. When a user logs on to Amazon, a cookie is sent to the computer [15]. By providing information to Amazon, the customer can get access to what Amazon calls *recommendations*. Using sophisticated software tools, Amazon can map a customer to a cluster, perform some mathematical calculations, and create a list of books or records that similarly interested customers have purchased. This form of data mining and affinity marketing is supposed to enrich a registered user's shopping experience as though a live person were quietly, unobtrusively accompanying the customer [8]. The more the user shops with Amazon, the better the user profile becomes [15].

86 of the sites surveyed by EPIC in 1999 used cookies, and two sites even would not let users visit without generating cookies. A test reveals that this also goes for a site like `www.hig.no`, while other sites like Finn.no (`www.finn.no`), Zett.no (`www.zett.no`), Amazon (`www.amazon.com`) and eBay (`www.ebay.com`) try to set a considerable number of cookies, but still allow you to browse the web site and search for items.

The default browser setting is to trust cookies within one domain. Netscape soon modified its browsers so that cookies from one site could not be given to another site [39]. This way, cookies can normally only be read by the web page that set it, but if other web pages within one domain tries to read it, they will also get access. This will not allow the use of cookies to correlate users' activities between many different web sites, to track the user's usage history and preferences. This could instead be done by adding cookies to GIF images that are served off third-party sites [39].

2.3.2 Profiling

According to Levine et al. [57], almost every web server keeps a record of every web page it serves, and it records the unique address of the computer to which it serves things. Every web page ever visited by a user may be recorded somewhere. Log files help address questions about the behaviour of visitors, including typical navigation sequences, referring site and time on site [59].

Sultan [73] discusses how consumer behaviour on the Internet has changed over the last years of the 20th century. According to this article, consumers are not willing to subscribe to Internet features to be able to use them. Consequently, companies often need to generate revenues from other companies advertising on their web site.

Some advertising companies, e.g. DoubleClick, integrate cookies in advertising banners to offer their advertising clients the most detailed information possible [72]. While displaying these advertisements, the advertisers place cookies on the browser. Since a single advertiser serves up advertisements to many unrelated Web sites that the user may visit, they can track the browser's movements across those sites and build a profile of the browsing pattern and preferences [43]. Statistical information is then supplied about users' surfing habits. Everytime an ad banner from DoubleClick is displayed, a cookie is saved. According to Schaeffer [72], DoubleClick displays over 53 billion banners every month. This creates a lot of cookies.

Other companies, such as Adfinity, combine these browsing patterns with personal information collected from other sources into fully identifiable profiles of the individual's online and offline behaviour [9]. According to Chesbro [15] and Arnold [8], this was also attempted by DoubleClick, by the company's purchase of Abacus Direct, a direct marketing company with an extensive database with names, addresses and other online shopper information. However, due to strong protests, among others from the Federal Trade Commission (FTC), the linking of information was postponed [15]. The use of profiling may represent significant benefits to consumers and marketers alike, or a major invasion of privacy, depending on the perspective. Perhaps the most contentious issue with regard to profiling is the lack of control over how the information is used and whether the user had an expectation that it might be used in certain ways [42].

According to the GetNetWise web page [43], some of the advertising companies have established a web site that allows users to opt-out of profiling or online preference marketing. DoubleClick is one of these companies [60]. If an opt-out option is chosen, a cookie will be stored on the user's computer to let the servers know that this browser has opted out in the past [42]. The irony is that this cookie will have to stay on the computer for as long as the user wants to be immune from the profiling, to tell the advertising company this every time they try to set a new cookie.

The WebEraser web page offers a program to erase Internet tracks left on the computer, like browsing history and selected cookies [79], which will also help protect against profiling. The web site of the Center for Democracy & Technology (CDT), has a page dedicated to online privacy. The web site talks about the combination of tools – legal, technical and self-regulatory – that are being designed to address the privacy concerns of Internet users [14]. Among the technical tools mentioned are P3P, proxies, anonymisers, cookies and system cleaners, all of which are also mentioned in this thesis.

2.3.3 Cache

In a typical web browsing session, a user may return to the same page several times. For example, a user may frequently return to a favourite home page, or repeatedly press the *back* button to retrace the steps through a site. Instead of incurring the bandwidth to re-request a page that has recently been viewed, modern browsers employ a local cache to store recently viewed content. Depending on the aggressiveness of the selected caching policy, i.e. how long the browser retains the cached page, a user can experience a marked improvement in anonymity. Once a site's pages have been retrieved and viewed, they may be very infrequently requested for re-transmission from the server. Consequently, data miners may be unable to construct a coherent view of the user's session, since the server is oblivious to much reading and browsing of its content [11]. Similarly, most browsers provide an offline browsing capability.

Another level of caching occurs at the corporate or ISP level [11]. ISPs normally cache often viewed web pages to make the loading of them faster. This caching could be done in any proxies used on the Internet, also in anonymising proxies. The solution for those who really want to use the Internet without 'anyone' knowing it (except Google administrators), is to only use the cache of Google (`http://www.google.com`). If the browser is modified not to request any images, only Google itself will be able to see the Internet use.

Web servers could counteract caching by using the 'pragma: no-cache' variable (see

Chapter 4.1.3 for demonstration). This would normally command all transporting instances not to cache the page. In these cases, the browser would reload pages from the server even when using the *back* button, and data miners would be able to trace the user's browsing pattern.

2.3.4 Sessions

Sessions, or session objects, are server-side collections of variables that make up the 'state', e.g. to tell that a user is already logged on to a service on the web page. The common approach to associate each set of data with the correct client is to have the client pass a session ID on each request. The most convenient way to make the client send the session ID on each request is to store it in a cookie as soon as the session is initiated [51]. Some systems choose to put the session ID in the URL.

Session hijacking is the process where an unauthorized user is able to get hold of the session ID of a logged in user. In this scenario there is no longer need for the user-name and password, as the session ID works as a 'short-time password' and tells the server that this user is already logged in. One of the methods used to find the session ID is packet sniffing. The only secure method to avoid session hijacking is to keep the session ID secret. One measure used against session hijacking is to tie the session ID to the IP address of the client [51]. This might create problems if the client is behind the proxy server of an ISP. In the case of just one proxy server, all the clients behind this server will have the same IP address. In the case where an ISP uses several proxy servers, one single client might send different IP addresses for each request, because of the load balancing performed by the proxy servers. The request is always sent through the proxy server which is least busy, which could lead to the user receiving a different IP address for each request [51].

2.3.5 Phishing

The new 21st century trend in online fraud is called 'phishing' [63]. This is a common term for all kinds of fraud where someone tries to trick or socially engineer an Internet user into giving away confidential information which can be misused [63]. The term is derived from fishing for persons who might be possible to con, e.g. by sending out millions of e-mails, and receiving few, but usable answers. Common for all types of phishing is that someone passes oneself off as a credible source, e.g. your bank or credit card company, and contacts you to ask you to confirm your credit card details and/or password [1].

According to Ollmann [63], the most successful phishing attacks have been initiated by e-mail, where the phisher impersonates the sending authority, for example by imitating the source e-mail address and embedding appropriate corporate logos. The victim receives an e-mail, supposedly from *support@mybank.com*, with the subject line 'security update', requesting them to follow an URL [63]. One way of tricking the user into thinking that they are sent to a web page which could be trusted is by creating a URL which looks like the real one, like shown in Figure 1 with the Norwegian newspaper Verdens Gang (VG, `http://www.vg.no`).

Another way of performing phishing is described by Allen and Hornberger [4, p. 131]. A person could post a message to an Internet bulletin board, using JavaScript to create a

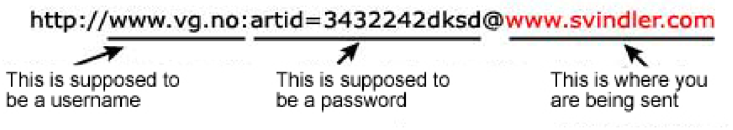

http://www.vg.no:artid=3432242dksd@www.svindler.com
This is supposed to be a username
This is supposed to be a password
This is where you are being sent

Figure 1: *Fake URL [76]*

new window and present the user with a form to log in. This form of social engineering could be used to trick unsuspecting users into entering their data and posting it to a completely foreign entity. As far as the user can tell, the HTML comes from a trusted source. JavaScript executes within the context of the page that initiated it. JavaScript thereby has the ability to harvest cookies, passwords from HTML forms, and users' browsing habits. Once the JavaScript runs, the damage is done. To avoid this problem it has become customary to validate and clean the data which is e.g. posted to bulletin boards [4]. One method of doing this could be by stripping out all HTML tags except *br* (linebreak), *ul* (unordered list), *li* (list item), *strong* (emphasised text) and *b* (bold text). This would prevent any malicious script tags from being executed, including the *meta refresh* possibility, which could be used in the HTML code. A user could be automatically redirected to another web page after a certain time, as shown in Figure 2.

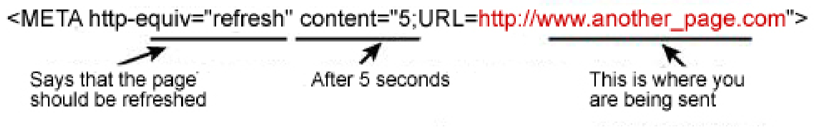
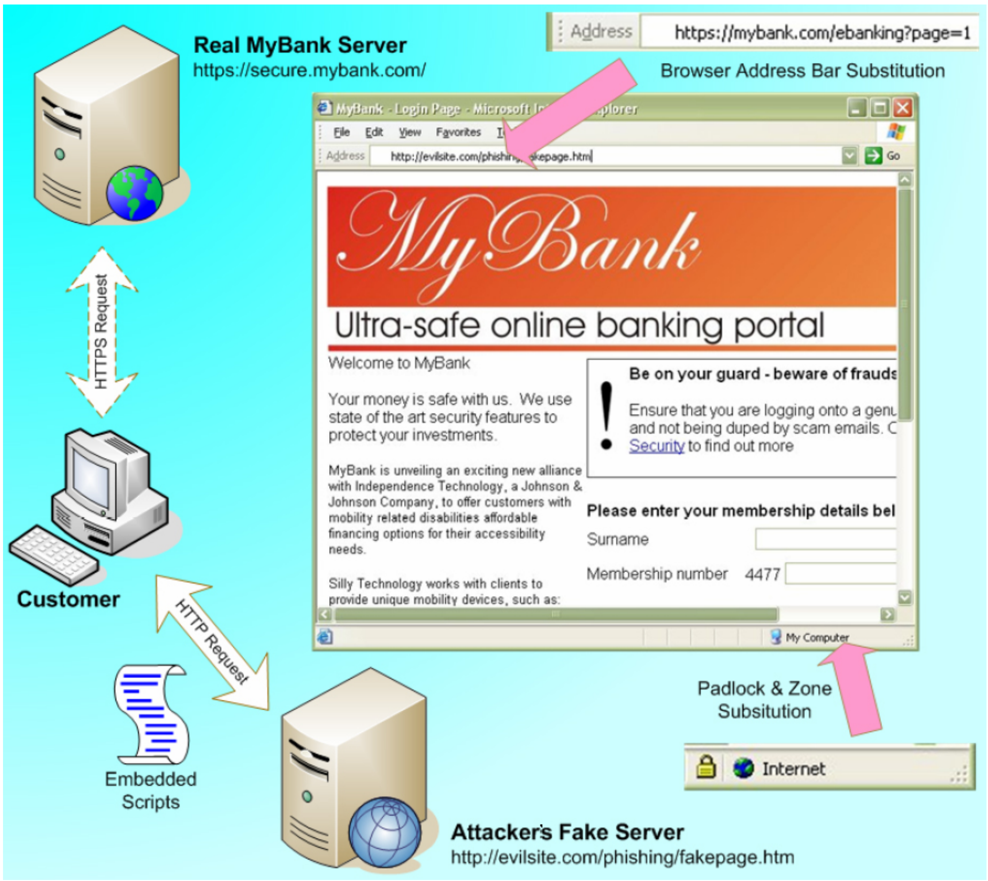

<META http-equiv="refresh" content="5;URL=http://www.another_page.com">
Says that the page should be refreshed
After 5 seconds
This is where you are being sent

Figure 2: *The META refresh tag*

The latest method that could be used in this kind of scam is fake certificates, which also will make the user think that the web page belongs to his or her bank or credit card company. Several methods exist for phishers to override displayed content [63]. While it is possible to override page content quite easily through multiple methods, one problem facing phishers is that of browser specific visual clues to the source of an attack. These clues include the address field, and the secure padlock representing an HTTPS encrypted connection, as well as the zone of a page source (see Figure 3 for illustration). Internet users have been taught that a padlock in the lower right corner means that the web page is approved and safe. The padlock used to be proof that a third party had confirmed the web page. A common method used to overcome these visual clues is through the use of scripting languages to position specially created graphics with fake information over these key areas [63]. Figure 3 shows how the attacker could use carefully positioned fake address bars and padlock/zone images to hide the real information [63].

Phishing reminds us of the *Nigerian Letter* scam, where the goal also is to lure as much information as possible from a person, to make oneself capable of extracting money from the victims' accounts. The Nigerian Letter scam started out in letter form, migrated to fax machines, and then ended up being transmitted by e-mail [58]. First appearing in the early 1980s, the original letter involved a request for help from an official of a government agency. This was often the central bank in Nigeria [18]. Today, the country


Real MyBank Server
https://secure.mybank.com/
Address
https://mybank.com/ebanking?page=1
Browser Address Bar Substitution
MyBank - Login Page - Microsoft Internet Explorer
File
Edit
View
Favorites
Address
http://evilsite.com/phishing/fakepage.htm
Go
HTTPS Request
Customer
HTTP Request
Embedded Scripts
My Computer
Padlock & Zone Subsitution
Internet
Attacker's Fake Server
http://evilsite.com/phishing/fakepage.htm

MyBank

Ultra-safe online banking portal

Welcome to MyBank

Your money is safe with us. We use state of the art security features to protect your investments.

MyBank is unveiling an exciting new alliance with Independence Technology, a Johnson & Johnson Company, to offer customers with mobility related disabilities affordable financing options for their accessibility needs.

Silly Technology works with clients to provide unique mobility devices, such as:

Be on your guard - beware of frauds

Ensure that you are logging onto a genu and not being duped by scam emails. C Security to find out more

Please enter your membership details bel

Surname

Membership number 4477

Figure 3: *Phishing by replacing graphics [63]*

in question could be any of a dozen, and the agency could use any official-sounding name.

The messages claim that the agency or bank in question has some really excessive funds, usually several million dollars, which need to leave country *X* for a while or it will be lost to the letter writer. Once the recipients show some interest, they are asked to pay for all kinds of things, e.g. advance fees and transfer taxes, and once the victim agrees to that first payment, there are many complications, all of which require still more payments [18]. The Nigerian Letter scam is also often called the *419 Fraud*, named after the relevant sections of the Nigerian criminal code [58]. The general rule for avoiding any con-game is: If something sounds too good to be true, then it *is* too good to be true [47].

Due to the phishers' high success rate, an extension to the classic phishing scam now includes the use of fake job sites or job offers. Applicants are enticed with the notion of making a lot of money for very little work – just creating a new bank account, taking the funds that have been transferred into it (less their personal commission) and sending it on as an international money order – classic money laundering techniques [63].

2.3.6 Pharming

A new threat to online transactions called 'pharming' has emerged. This threat illegally redirects users to fraudulent web sites [5]. This latest form of attack redirects Internet users from legitimate Web sites to malicious ones using a strategy called DNS cache poisoning. Although DNS cache poisoning is not new, the complexity of the new pharming attacks is cause for concern.

The 'Pharmer' inconspicuously hijacks the computer and coerces it into taking the user to a copycat web site. The site is most commonly a page that looks identical to that of a bank or financial institution. From this point, the user is tricked into submitting the passwords and financial information straight into the pharmers' data banks. The process can be compared to switching a street sign on drivers in a new city, sending them down the wrong street. Similarly it can also be compared to switching the names connected to phone numbers in a phone book, when a user goes to look up a name, they end up calling the wrong number [5].

When a user types a URL such as `www.google.com` into their Internet browser, a request goes to a local DNS server, which then locates the registered IP address for the Web server. When malicious users poison a DNS server, they change the IP address for a domain and send visitors to a completely different Web site, usually without the visitors' knowledge.

According to Anonymizer [5], pharming scams take several forms:

- A hacker could break into an Internet service provider's DNS servers and switch legitimate addresses stored in the server's 'cache' with bogus addresses (DNS poisoning).
- A scam artist could pretend to be a Web site's operator to persuade an Internet registrar to make the change to the bogus address in the registration database.
- Attackers could use malicious code, such as a virus or Trojan program planted on a user's PC, to track keystrokes or change a computer's settings to take users to fraudulent copies of legitimate Web sites they request.
- Hackers could also target the 13 'root' DNS servers that route all Internet traffic.

One way to check to see if the site you have been directed to is real is to look for the lock icon, located in the bottom right corner of the browser. If the icon does appear, one ought to click on it to verify a secure connection. However, as mentioned earlier, the lock does *not guarantee* security. DNSSEC was designed to protect the Internet from certain attacks, such as DNS cache poisoning [30]. It is a set of extensions to DNS, which provide origin authentication of DNS data, data integrity, and authenticated denial of existence. By checking the signature, a DNS resolver is able to check if the information is identical (correct and complete) to the information on the authoritative DNS server [30].

2.3.7 HTTPS

When commercial sources boast about 'secure web servers' they normally refer to web servers capable of doing encrypted communication [51]. In a web setting, encryption usually means HTTPS. This may be described as HTTP communication over an encrypted channel. The encrypted channel is provided by SSL [51]. The encryption protects the network connection between the client and the server. If everything works as expected, HTTPS makes it impossible for someone to listen to traffic in order to extract secrets. People may still sniff packets, but the packets contain seemingly random data [51]. However, they may still see who is talking to whom, as HTTPS only protects the content of the traffic, not the origin.

2.4 Protecting privacy

There are a lot of different ways to protect a user's privacy when browsing the Internet. For example, several privacy seal programmes exist, for example TRUSTe, BBBOnline and WebTrust. However, the seal programmes merely verify that a Web site's privacy policy discusses certain privacy topics, like the use of cookies and sharing data with third-party marketers [57]. The seals do not set any specific quality standards, benchmarks or practices. However, by displaying one of these online privacy emblems on its website, a company at least communicates to its users its commitment to follow the tenets of the particular programme.

If an organisation posts a policy and then does not verify that the policy is being followed in actual practice, the privacy policy can become more of a liability than an asset [19]. If a US company has posted a privacy policy and then violates the policy, the company can be prosecuted for fraud and deceptive trade practices. Ironically, if a company *does not* post a privacy policy, it has made no public promises to which it can be held, and the company cannot be touched [57].

2.4.1 Anonymising technologies

In 1997, *anonymous remailers* for electronic mail were an established technology [45]. At first, the servers just stripped the headers and resent the message [46]. The first well-known remailer was *anon.penet.fi*, providing anonymous and pseudonymous e-mail account as early as 1993 [24]. Gradually, the technologies included cryptography and fixed-sized packages to add extra security to the data as well as anonymity to the sender (*Mixmaster*, or type II, remailers) [45]. The most widely used protocols for remailing are Mixminion and Mixmaster. The 'strip identity headers and resend' approach used by remailers was also applied to provide anonymity protection for web browsing as well [46]. Perhaps the most promising upcoming technology in 1997 was Wei Dai's proposal for *PipeNet*, a service analogous to the remailer network, but designed to provide anonymity protections for real-time communication, such as web traffic, interactive chats and remote login sessions [45].

Unfortunately, PipeNet was never developed past the initial design stages. In 1998, Demuth and Rieke [26] presented a system called *Janus*, where both the server and the client could remain anonymous. To maintan the client anonymity, the Janus system filtered the header fields, and replaced the information with its own information, or removed it completely. The Janus system was later known as *the Rewebber* [27].

One available anonymity network to enable anonymous web browsing is the MIX network. It offers *sender anonymity* and *relationship anonymity* [67]. The core of the MIX network consists of *anonymity proxies*. The proxies are distributed over the Internet, and play the role of the MIXes in the system. The goal of the proxies is to hide the correlation of incoming and outgoing messages, such that an attacker cannot follow a message through the network [67].

A MIX is a special network station which collects and stores incoming messages, discards repeats, changes the appearance of the messages by encryption, and outputs them in a different order. By using one MIX, the relation between sender and recipient is hidden from everybody but the MIX and the sender of the message [39].

Anonymizer is a system corresponding to one single mix, to provide anonymity protection for web browsing. It is a web proxy that filters out identifying headers and source addresses from the web browser. The connection itself is however not anonymised [39]. Anonymizer consists of a single point of trust with rather weak security features. A more secure solution is provided by *onion routing* [39].

Onion Routing was starting to get deployed in 1997, and attempted to accomplish similar goals as PipeNet. Onion Routing, however, chose to trade off more towards performance and robustness. In contrast, PipeNet chose security and privacy above all else, to the extent that it preferred to shut down the entire network if the alternative was to leak a bit of private information [45].

MIX networks are not suitable for quick interaction like web browsing and web chats [67]. A basic anonymising proxy like *Anonymizer* could instead be used to protect the user's identity on the Internet. Traffic appears to come from the proxy instead of the user. However, this single proxy would be vulnerable to faults, attacks, traffic analysis, etc. Onion Routing combines the advantages of MIXes and proxies, by encrypting the packages in addition to hiding the user's identity [67]. Onion Routing is a distributed overlay network designed to anonymise TCP-based applications like web browsing, secure shell and instant messaging [29]. Clients choose a path through the network and build a circuit, in which each node ('onion router') in the path knows its predecessor and successor, but no other nodes in the circuit [29].

Tor is a network of virtual tunnels built on the onion routing principle, that allows people and groups to improve their privacy and security on the Internet. It also enables software developers to create new communication tools with built-in privacy features. Tor provides the foundation for a range of applications that allow organizations and individuals to share information over public networks without compromising their privacy [29].

The idea of Tor is similar to using a hard-to-follow route in order to throw off somebody who is tailing the users – and then periodically erasing their footprints. Instead of taking a direct route from source to destination, data packets on the Tor network take a random pathway through several servers that cover the tracks, to make it impossible for any observer at any single point to tell where the data came from or where it is going [29]. Communications are bounced around a distributed network of servers, the onion routers.

2.4.2 The Future

The rapid growth of the Internet has virtually depleted the available inventory of 32-bit IP numbers. Accordingly, the IETF has developed a specification (IPv6) for a new generation of 128-bit IP numbers [11]. Using 'stateless address auto-configuration', an Internet user's computer would automatically generate its own IP address. The IPv6 privacy extensions introduce randomisation into the address auto-configuration process, so that the resulting address cannot be directly asociated with the originating hardware. In addition, the extension specifies that at least once a day a new IPv6 address is randomly generated, and in a manner that the new address cannot be correlated with previous numbers [11].

2.5 Regulation by law

The Internet exists within social, political, and technological contexts that can impede its democratic potential. Governments worry about the Internet's threat to their traditional authority [9]. Recent years have seen an escalating trend in various jurisdictions to codify privacy rules into local law [45].

The very first modern data protection act was adopted by the Parliament of the West German state Hesse in 1970, and served as a basis for the adoption of similar laws by other German states and the West German federal government, as well as governments outside Germany. Sweden's Data Act, which was passed in 1973, was the first national data protection act in the world. These data protection acts were followed in Europe by laws enacted in France, Austria, Denmark and Norway in 1978 and in Luxembourg in 1979. In the USA, the Privacy Act was adopted by the Congress in 1974 [39].

According to Clayton and Stewart [19], a new bill was introduced to the US Congress in 2002, to help protect US consumer privacy on the Internet. This bill has however at the time of writing this thesis still not been passed. Privacy has until now been protected by the old Federal Trade Commission Act. This Act does not include any requirement that a web site must have a privacy policy. It requires only that once a company has a web privacy policy it must abide by it [19]. The proposed bill would make companies and other groups disclose somewhere on their web sites what they do with customers' personal data.

2.5.1 A borderless community

On the Internet, information and communications flow unimpeded across national borders. National laws may be insufficient on their own to provide citizens with privacy protection across borders [9]. An important problem is deciding which laws should govern, for example when a server residing in one country is storing information about a citizen of another country. Directive 94/46/EC, adopted in 1995 by the European Union (EU), is the most comprehensive and complex of the legal instruments on data protection which have been introduced at the international plane and provincial and municipal levels [12]. It constitutes also the most important point of departure for new data protection initiatives, both in and outside the EU. The Directive exercises some political and legal influence over other countries outside the EU, not least because it prohibits, with some qualifications, the transfer of personal data from the member states and the EEA (Norway, Iceland and Liechtenstein) to these countries unless they provide 'adequate' levels of data protection [12].

Directive 94/46/EC forced the United States government and industry leaders to carefully re-examine the US privacy policies [14]. The proposed international safe harbour principles were approved in 2000, to enable corporations to run multinational operations and meet the EU standard for adequate privacy protection. The safe harbour privacy principles are self-regulatory privacy guidelines that shall prevent US companies' data transfer from being cut off by the EU [39]. In July 2002, the EU adopted a new directive (Directive 2002/58/EC) translating the principles of Directive 94/46/EC into specific rules for telecommunications and other electronic communications, addressing privacy and security, marketing, cookies, data retention, and so on [14].

Other instruments that ought to be mentioned are the CoE Convention, the OECD Guidelines, and the UN Guidelines. The different guidelines and conventions differ on whether they value the *free flow of data across national borders* or the *restriction of data flow to countries without equivalent or adequate levels of data protection* [12]. The OECD Guidelines have influenced international agreements, national laws and self-regulatory policies [14]. The goal of the OECD Guidelines is to ensure that consumers are just as safe when shopping online as off-line, no matter where they live or where the company they do business with is based [8].

In a lawsuit in 2000 the New York courts ruled against an Antigua casino that they had violated both state and federal gambling laws [18]. According to the verdict, it is irrelevant that Internet gambling is legal in Antigua. The act of entering the bet and transmitting the information from New York via the Internet is adequate to constitute gambling activity within New York State. In another ruling, a French court required in November 2000 that Yahoo! (`http://www.yahoo.com`) blocks access in France to US based auction sites selling Nazi memorabilia and artefacts [18]. The problem is how to enforce these rulings.

2.5.2 What is protected

According to Bygrave [12], data protection laws can be described as laws comprised of rules that specifically regulate all or most stages in the processing of certain kinds of information. Only *personal* information is usually covered by data protection laws. Such information is typically defined in these laws as information relating to, and permitting identification of, individual physical/natural persons (individuals) or sometimes collective entities. This implies that information identifying the user is regulated, while information about which pages the user has visited is not regulated. Data linked to machines and other non-human objects, like IP addresses and user-names, could elude regulation pursuant to most data protection laws [12]. However, if the latter information is linked to a specific user, the data protection laws will usually come into effect.

According to Directive 94/46/EC a computer's IP number could qualify as an identification number *relating to an identifiable natural person* [12]. Recital 26 in the Directive's preamble indicates that data will only be personal if they can facilitate identification of a single person by means that are reasonably capable of being utilised by another person. In the event where a readily accessible directory exists, listing one particular person against one particular IP address, there will be a relatively high chance of that address (and the other click-stream data registered against that address) constituting personal data. The chance will be lessened in cases where the Internet service provider issues a temporary address and fails to keep a record of which user name has been registered against that address [12].

Crafting proper privacy protections in the electronic realm has always been a complex endeavour. It requires a keen awareness of not only changes in technology, but also changes in how the technology is used by individuals, and how those changes are pushing at the edges of existing laws [9].

2.5.3 Government use

Government is one of the biggest consumers and producers of dossiers of personal information, and as such could be viewed as a potential threat to privacy [46]. In the hands of government, profiling and cross-referencing profiles with behaviour patterns could help catch criminals, or attempt to predict future criminal behaviour [42].

Carnivore and *Magic Lantern* are technical surveillance technologies used by the US Federal Bureau of Investigation (FBI) [42]. The FBI has priorities that range from terrorism, espionage protection and cyber attacks to more traditional general crime fighting roles. Also on the list of the FBI's priorities is to *protect civil rights*. This role may seem in conflict with the FBI use of powerful civilian surveillance technologies [42]. The Carnivore software allows capture of e-mail. Magic Lantern is a 'key logging' software. Magic Lantern must be installed on the target machine to work, and is therefore similar to commercially available spyware [42].

Passed in the wake of the September 11 2001 terrorist attacks on the World Trade Center, the US Patriot Act broadened a variety of law enforcement privileges. Some of its provisions affected existing law regarding privacy, or otherwise impacted perceived privacy issues, such as the ability to intercept communications for domestic and foreign intelligence gathering [42]. There have been numerous instances of local authorities questioning the breadth, depth and whole appropriateness of all the Act's provisions. The Patriot Act may have been hastily passed in an emotionally charged environment, and it may be in the national interest for there to be expansion of law enforcement capabilities. However, the scope of change in the Patriot Act and the vagueness of some of the wording make these regulations highly suspect, not merely on a *right* or *wrong* level, but even in that they create a great deal of ambiguity with respect to long existing law and doctrine [42].

2.5.4 Focus

Over the past decade, experience has shown that the best way for Internet users to protect their privacy is self-help – learning about the dangers of the Internet, their rights, and the available privacy tools [48]. Conflicting state laws often just lead to more user confusion.

Together, the characteristics of the Internet medium pose challenges to the traditional top-down methods of implementing policy and controlling behaviour. Providing a seamless web of privacy protection to data as it flows through this international network will, according to Berman and Mulligan [9], require us to harness the business community's interest in promoting commerce, the government's interest in fostering economic growth and protecting its citizens, and the self-interest of individuals in protecting themselves from the overreaching of the government and the private sectors. It requires us to use all of the tools at our disposal – international agreements, legislation, public education and the technology itself. Berman and Mulligan [9] claim that we must begin by reaching consensus on what we mean by protecting privacy, but we must keep the characteristics of the online environment sharply in focus.

In a fully networked society, privacy is seriously endangered and cannot be sufficiently protected by privacy legislation or privacy codes of conduct alone. Data protection commissioners are therefore demanding that privacy requirements should also be technically enforced, and that privacy should be a design criterion for information systems [39]. Some of these privacy technologies are in the hands of the individual users, who can decide to use them to protect themselves. Other privacy enhancing technologies are implemented by a privacy-enhanced system design, for instance by avoiding or minimising the collection and use of personal data [39].

Generally, privacy education is important to raise the awareness of the users. Most privacy-enhancing technologies by themselves are not necessarily an effective means to technically enforce privacy aspects, unless users have sufficient technical knowledge to apply them. Thus users need information and education about their rights, about the value of their personal data, about privacy risks and the possibilities of self-protection [39].

3 Web page experiments

A web page was created to demonstrate what kind of information it is possible to extract about a visitor. An experiment was performed, to find out what kind of information it is possible to store about the visitors. Through this process the thesis has established the level of knowledge necessary to create such a web page and to collect the wanted information.

A survey was originally planned performed, to see whether people trust the Internet as a medium. This was found unnecessary, as this kind of survey has been performed both by TNS Norsk Gallup [40] and the European Commission [33]. This gave as good, and possibly even better, results than a survey performed during the thesis work, and the results are summarised in Chapter 2.2.2.

3.1 Ethics

It is not recommended to make an open web page collecting information about casual visitors. Collecting information on Internet users without their knowledge could be a breach of ethics. It was important that the panel participating in the laboratory test was fully advised. They received information on what kind of research was being done, and what information would be stored, as well as what the information would be used for. The invitation to participate in the experiment, sent by e-mail, is added in Appendix B.

Because the object of the experiment was to find out what kind of information can be revealed and stored, it was important that the subjects were fully informed and still wanted to participate. It was also important that people who were not informed were unable to get access to the web site. The URL was therefore chosen to be one which is not so easy to guess, `http://studweb.hig.no/001635/MSc`. Participants were informed through an invitation by e-mail, and the information was repeated on the web page. When entering the web page, participants would have to press an *OK* button to say that they wanted to participate. One problem is that it is not possible to control that the participants have read the information thoroughly and understood it, and not just pressed the *OK* button.

To avoid ethical dilemmas, the registration was only of *what kind of* information was revealed, the data itself was only displayed to the participants, and never stored. This gave a sufficient statistical foundation for the further work of the thesis, as the important part was to demonstrate which information can be stored, and how easy it is to store it.

3.2 Theory

A web page was made for the experiment, to be able to decide how difficult it would be to extract the information from the visitors' browsers. It was also important to see what kind of information could be extracted. The web page of the laboratory test was later used as a metric when testing the level of protection given by the privacy protection software. To measure the amount of information revealed, the web page was used to see what parts of the available information was revealed. It was important to test different browsers, e.g. Microsoft Internet Explorer, Netscape Navigator, Mozilla Firefox, Opera and Konqueror. Whether a firewall software can protect against information getting revealed was also tested.

The web pages were coded using PHP with HTML and JavaScript. PHP was chosen because it has a lot of functions built in to extract information from browsers. Other technologies could also be used, for example Java Applets. However, this cannot find other information than can be found by PHP and JavaScript. In particular, applets cannot find out any information about the local computer except for the Java version used and the name and version of the operating system [49, p. 592]. This is because applets are interpreted by the Java Virtual Machine in the browser, and not directly executed by the user's computer.

3.2.1 PHP

HTML is a client-side technology, meaning that an HTML document is processed entirely by the client. A web server merely provides requested files, the client browser makes the decisions about rendering them. PHP, conversely, is entirely server-side. When a PHP script executes, it does not interact directly with the browser. Only the final product of the PHP script, which is usually an HTML document, is dealt with by the requesting browser. Browsers cannot execute PHP scripts. Figure 4 illustrates a client request for a PHP script (right), compared to a request for an HTML document (left). Before the document is sent to the client (the browser), the document is processed by the PHP engine, which executes any PHP code found in the document. The PHP script in figure 4 returns a processed HTML document. The same approach is used with other server-side technologies, e.g. JSP and ASP.

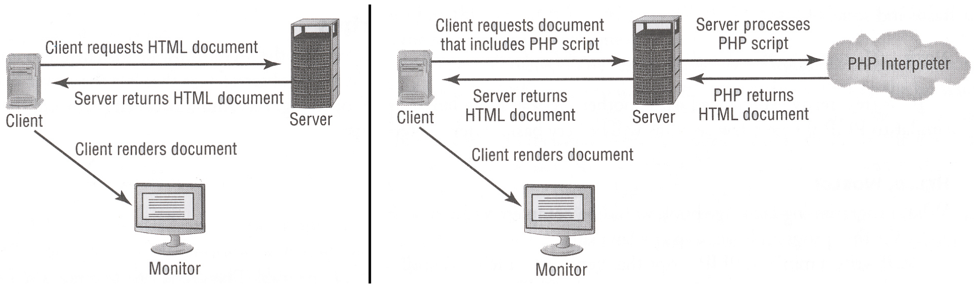

Client requests HTML document
Server returns HTML document
Client
Server
Client renders document
Monitor
Client requests document that includes PHP script
Server processes PHP script
PHP Interpreter
Server returns HTML document
PHP returns HTML document
Client
Server
Client renders document
Monitor

Figure 4: *HTML vs. PHP script request [4, p.5]*

3.2.2 JavaScript

JavaScript is a client side scripting language that can be used to integrate small programs in a web site. The programs are executed when the web site is viewed [72]. JavaScript should according to Schaeffer [72] be secure and not allow any destructive actions and attacks. However, security holes in web browsers are easy to exploit. For example the address in the browser address line could be falsified, and visitors could be led to believe that they are looking at the site of another supplier while the hacker obtains accessing data. A lot of weaknesses of previous browser versions have been eliminated, and current browsers are not so liberal with JavaScript [72]. JavaScript is also often used to send information about a visitor back to a web server [47].

3.3 Experiment

A web page demonstration was performed, to demonstrate what kind of information it was possible to find about visitors to a web page. It was also important to find the difficulty of extracting the information, and what kind of knowledge was necessary to create the web page. PHP and JavaScript manuals were studied, as well as discussion forums on the Internet, to find as many exploits as possible.

3.3.1 Web page

The experiment was performed with invited participants. 40 participants received an invitation per e-mail, and 69 test participations were registered (some participants tried several times with different computers). The text in the e-mail explained the purpose of the experiment, and what information would be stored. The e-mail is, as mentioned earlier, reproduced in Appendix B.

When the participants entered the web page, the information about the experiment was repeated, to make sure the participants were fully informed on the experiment, and on what information was going to be stored for further use. The first page viewed by the participants is shown in figure 14 (Appendix B). The participants were also asked to select which firewall software was used, if any, to make it possible to see if there was any difference in which information was displayed.

Figure 15 (Appendix B) shows the results web page, as viewed in Mozilla Firefox and most other browsers. The information that could have been collected is displayed to the user, as well as what limited information was registered for the experiment.

In figure 16 (Appendix B), the difference when using Internet Explorer is shown. While most browsers tend to ignore the script collecting the clipboard content from the computer altogether, Internet Explorer displays it.

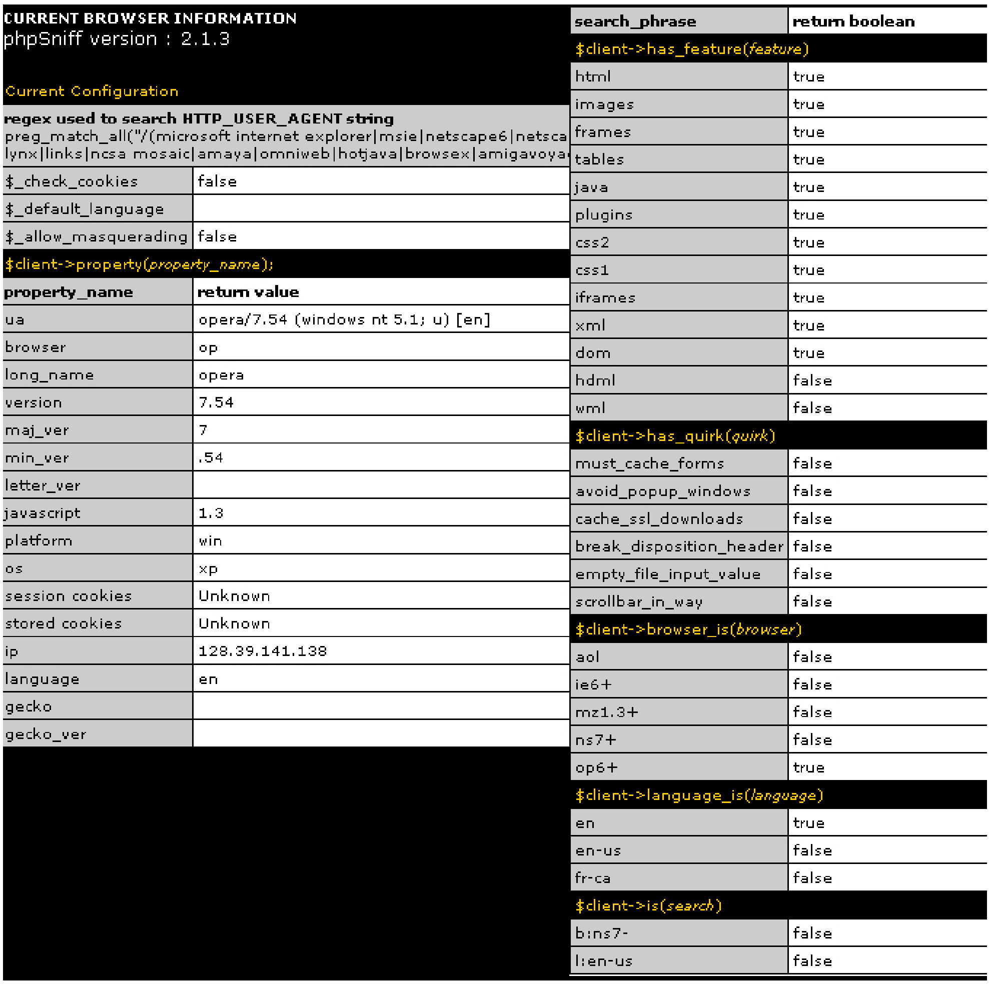
The information had to be gathered somehow. Several solutions were possible, for example sending data to a database or a text document. Another possibility was to just send the information via e-mail to the author. As there was not going to be a large participation in the experiment, a database would have created unnecessary work, as the number of entries would be limited. The information was therefore sent to the author by e-mail, for manual registration.

3.3.2 Source code

Most of the code used in the experiment is demonstrated in several PHP tutorials, for example W3Schools [78] and books like Gilmore [44] and Ullman [75], and is accordingly something anyone can find and use. Servers may hold information such as which URL the user came from, the user's browser, and other information. This information is stored in variables. These variables could be retrieved by calling server variables by the PHP code *$_SERVER["server-variable"]*. The server-variable *HTTP_REFERER* stores the web site from where the client was referred (the Referer header), while the server-variable *HTTP_USER_AGENT* will give the name and version of the client's browser.

REMOTE_ADDR stores the IP address of the client. The following code will look up the IP address, find the name of the computer at that address if the address exists, and store it into a variable: *$host = gethostbyaddr($REMOTE_ADDR);*.

According to Ullman [75], another option for browser detection is to use a class such as phpSniff, available at `http://sourceforge.net/projects/phpsniff`. The class, as it stands, is used to detect almost anything one would want to know about the visitor's browser (Figure 5). This class was selected for the experiment, as it gave a more detailed browser information than the server variables.



CURRENT BROWSER INFORMATION
phpSniff version : 2.1.3

Current Configuration

regex used to search HTTP_USER_AGENT string
preg_match_all("/(microsoft internet explorer|msie|netscape6|netsca
lynx|links|ncsa mosaic|amaya|omniweb|hotjava|browsex|amigavoya

$_check_cookies	false
$_default_language	
$_allow_masquerading	false

$client->property(*property_name*);

property_name	**return value**
ua	opera/7.54 (windows nt 5.1; u) [en]
browser	op
long_name	opera
version	7.54
maj_ver	7
min_ver	.54
letter_ver	
javascript	1.3
platform	win
os	xp
session cookies	Unknown
stored cookies	Unknown
ip	128.39.141.138
language	en
gecko	
gecko_ver	

search_phrase	**return boolean**
$client->has_feature(*feature*)	
html	true
images	true
frames	true
tables	true
java	true
plugins	true
css2	true
css1	true
iframes	true
xml	true
dom	true
hdml	false
wml	false
$client->has_quirk(*quirk*)	
must_cache_forms	false
avoid_popup_windows	false
cache_ssl_downloads	false
break_disposition_header	false
empty_file_input_value	false
scrollbar_in_way	false
$client->browser_is(*browser*)	
aol	false
ie6+	false
mz1.3+	false
ns7+	false
op6+	true
$client->language_is(*language*)	
en	true
en-us	false
fr-ca	false
$client->is(*search*)	
b:ns7-	false
l:en-us	false

Figure 5: *phpSniff demonstration [75, p.280]*

As soon as the phpSniff class was initiated, the wanted properties were received by calling them from phpSniff with the code *$client›property('property-name');*. The property-names are shown in figure 5, e.g. *ip* (ip address), *ua* (user agent / browser information), *os* (operating system) and *javascript* (JavaScript version installed). The *long_name* and *version* properties were supposed to give the browser name and browser version, but it seemed to have problems identifying Microsoft Internet Explorer, as this uses a different format in the variable shown here as *ua*, placing browser name and version in a different position from what other browsers do (see figure 16, Appendix B, for an example of this).

The *setcookie()* function allows you to set a cookie on a user's browser. The function comes in two basic types [20], session cookies, and cookies that expire after a certain amount of time [50]. To set a cookie, the JavaScript code is: *setCookie("name", value, expiration);*. The cookie possibility was used in the laboratory test to send the information of how many pages the participant had visited before entering the test. The code to extract the number of pages is the simple JavaScript *history.length* [53].

The JavaScript code *clipboardData.getData("Text");* will collect the data from the windows clipboard [68]. The script was found by a simple search on the Internet. The web page which describes this script also explains how to forward the content to the web server or another user, and claims the reason for this to be that they want Microsoft to improve the security of the Internet Explorer browser.

There is also a possibility to locate the visitor geographically by way of the IP address. According to Kabir [54, p. 583], the netgeo.php class is able to trace route where a certain IP address is located geographically. The class was tested and found somewhat inaccurate, as it placed many Norwegian IP addresses in the wrong geographic location, e.g. the Netherlands. Another geographic locator tool, this one a script from a company called GeoBytes [41], was found to be more accurate, and was therefore used in the laboratory test. With the help of this tool, the visitors' geographic location was displayed to them.

3.4 Results

Both the Anonymizer website [5] and the FindNot website [38], as well as several others, present demonstrations on how easy it is to find information about a user. This was confirmed through this experiment, which proved that extracting information about visitors is fairly easy. The results of the laboratory test is shown in table format in the last part of Appendix B. It was not possible to see any differences between the browsers or the firewalls that were registered.

For this reason, the laboratory test turned out to give no usable statistics for the further work in this thesis. However, it created interest in the thesis work and made people aware of what kind of information it is possible to collect. The only statistical information possible to extract was the fact that if cookies were not allowed, the number of pages visited was not shown, due to the use of cookies to store this information in the test.

However, the web page made for the laboratory test was still needed for the software testing in the next part of the thesis.

Some of the server variables shown in this experiment are set by the client, and one way of protecting against this information being revealed, is replacing the information with fake information. The Referer header and user agent variable could for example be set by the clients themselves. On the other hand, the variable which collects the IP address is usually correct, and cannot be easily changed by the client. It can however be changed by a proxy. When the IP address is changed, the name of the computer can normally also be changed, as well as geographic location.

It seems not to be very easy to find information on how to use a web page to extract information about visitors, even though most tutorials on web programming describe simple scripts to find information such as IP addresses, browser versions and operating systems. This makes the whole process of storing information about visitors to a web page more difficult, except when using other elements such as cookies to trace browsing patterns. It seems like except the IP address and the browser information, it is necessary to do a bit more research to be able to find out how to get the information. The script to extract information from the clipboard and to find the geographic location was for example not that easy to locate. However, if a person has decided to e.g. exploit the clipboard, the information on how to do it is out there, and can be found by doing a simple search using search engines such as Google.

The script to find whether or not a proxy is used seemed not to work in the experiment. Normally the variables to decide if a proxy is present are set by the proxy itself, so consequently the proxy decides whether it will disclose its presence or not. None of the participants in the experiment seemed to be using a proxy. This is probably because if proxies are used, they could be transparent or set up not to reveal that they are present, and consequently are not revealing their existence. This was tested by using proxies on the test computer, and the proxies seldom revealed themselves.

There is a slight difference between browsers regarding what kind of exploits they will allow. The main difference shown in this experiment is the clipboard exploit, which only works in Microsoft Internet Explorer. The other browsers tested (Mozilla Firefox and Opera) will only disregard this script in total. Revealing the clipboard information makes it possible to pass this information on to someone who might be interested in collecting this. This was one of the most threatening findings from this experiment, leaving an adversary with the possibility to wait for passwords, account numbers etc. being copied to the clipboard.

Earlier versions of browsers has also allowed exploits of information like the URL of previous visited pages. According to several JavaScript tutorials, for example JanetSystems [53], the browser history should be possible to access e.g. by the code *history.previous*. This possibility seemed not to work in current browser versions. Several sources, for example Gralla [47], Levine et al. [57] and Demuth and Rieke [26] also claim that the browser can reveal the e-mail address of the user, as set in the default mail program. This has not been verified in this experiment, and is also not found in recent web scripting tutorials, something which might imply that this information is also obsolete.

4 Privacy tests

The following part of the thesis aims to find ways to protect the user. Through experiments and literature studies, programs which claim to protect our privacy have been surveyed and tested to see whether they really do protect the user. A literature study has also been performed, to find out if there is a detection system available to help detect information being stored when visiting a web page.

4.1 Information detection

It is not easy to find information on how to detect attacks on an Internet user's privacy by way of a web page. When a typical browser fetches a web page, it issues an HTTP *GET* request to the address indicated by the page's URL and receives in response an HTML-object which may in turn contain references to other web objects [74], e.g. graphics and links to other pages.

To be able to see if it is possible to detect the different information being gathered via a user's browser, a Man-in-the-middle proxy was used, to see which information is visible when browsing the Internet.

4.1.1 Man-in-the-middle proxy

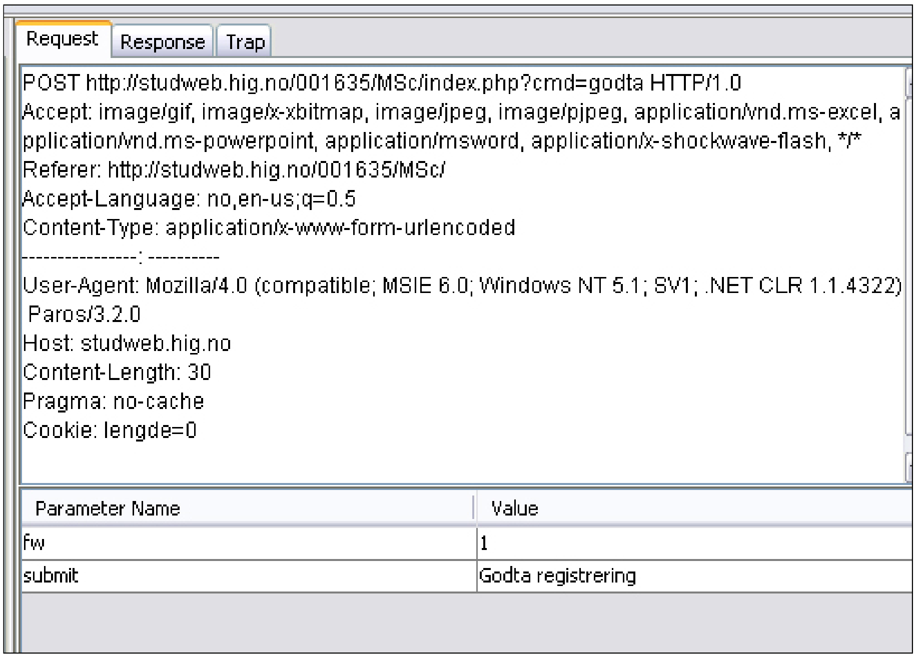
A typical HTTP proxy will relay packets to and from a client browser and a web server. A man-in-the-middle proxy will intercept an HTTP session's data in either direction and give the user the ability to alter the data before transmission [64].

4.1.2 Paros

Paros was chosen as the test proxy for the information detection experiment [16]. Paros is a proxy normally used for evaluation of the security of web applications. It is free of charge and written completely in Java. Through Paros' proxy nature, all HTTP and HTTPS data between server and client, including cookies and form fields, can be intercepted and modified [16]. For example, during a normal HTTPS connection a typical proxy will relay the session between the server and the client and allow the two end nodes to negotiate SSL. In contrast a man-in-the-middle proxy will pretend to be the server and negotiate two SSL sessions, one with the client browser and another with the web server. As data is transmitted between the two nodes, the proxy decrypts the data and gives the user the ability to alter and/or log the data in clear text before transmission [64]. However, the browser will in the case of an HTTPS session detect that the certificate name is invalid. In normal HTTP sessions this is not an issue, and the user will not be given a warning.

4.1.3 Log

The log from Paros, here displaying the *POST* command sent when pressing the *OK* button to accept participation in the web page experiment, is shown in Figure 6. The server variables and cookies are also displayed. All of the information displayed alterable, including the *User-Agent* information and the *fw* (firewall) variable. It is, however, not possible to see whether the server variables are stored or registered anywhere, only that they are sent.



```
POST http://studweb.hig.no/001635/MSc/index.php?cmd=godta HTTP/1.0
Accept: image/gif, image/x-xbitmap, image/jpeg, image/pjpeg, application/vnd.ms-excel, a
pplication/vnd.ms-powerpoint, application/msword, application/x-shockwave-flash, */*
Referer: http://studweb.hig.no/001635/MSc/
Accept-Language: no,en-us;q=0.5
Content-Type: application/x-www-form-urlencoded
----------------:----------
User-Agent: Mozilla/4.0 (compatible; MSIE 6.0; Windows NT 5.1; SV1; .NET CLR 1.1.4322)
 Paros/3.2.0
Host: studweb.hig.no
Content-Length: 30
Pragma: no-cache
Cookie: lengde=0
```

Parameter Name	Value
fw	1
submit	Godta registrering

Figure 6: *Paros log*

The instruction *Pragma: no-cache* is a command to all transporting instances not to cache the corresponding page. In particular, the web browser of the user does not cache the page into a file on the computer's hard drive, like it normally does [27]. The no-cache function is especially important when sending forms on web pages. However, a proxy might still ignore this instruction.

The proxy makes it possible to see if the server sets a cookie, like here, the cookie with name *lengde*. The variable of the cookie, also named *lengde*, is set to '0'(zero), meaning no pages have been visited previously in this browser window. It is also possible to see all variables sent from a form, in this test the variables *fw* and *submit*. The variables from the test web page are displayed as *fw=1&submit=Godta+registrering*. There is also an option to display variables in table format, to make it easier to separate them, as shown in figure 6.

4.2 Privacy software

If someone wants to browse the Internet anonymously, the goal is to mask the IP address and as much other information as possible from the servers visited [42]. 'Anonymisers' are tools and services designed to help individuals surf the web or send email anonymously. These tools focus on minimising the risk that web requests can be linked to an IP address from which a user can be identified. Most of these services rely on a trusted third party (a proxy) which strips off identifying information like the IP address, and forwards requests on behalf of a user [21]. A variety of these anonymous proxy services are available both as free and fee based services. Figure 7 shows a diagram of the Internet traffic going through a proxy with Anonymizer Anonymous Surfing.

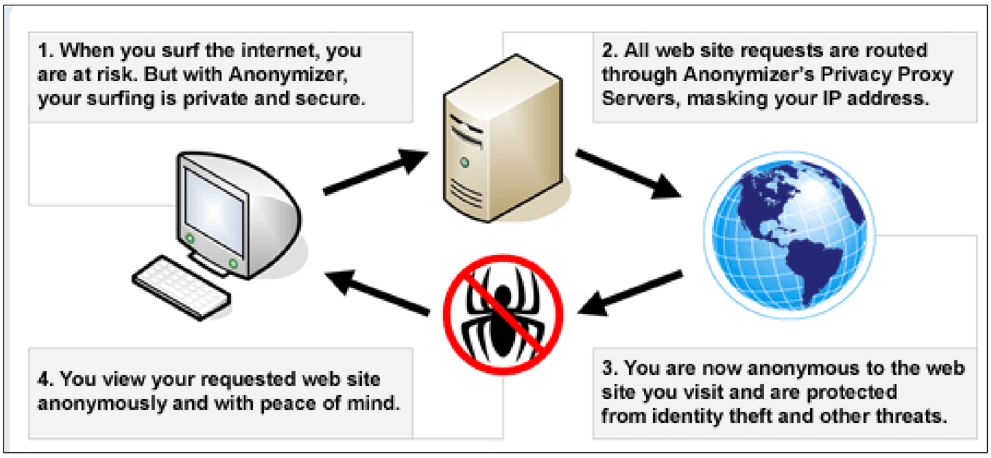

1. When you surf the internet, you are at risk. But with Anonymizer, your surfing is private and secure.
2. All web site requests are routed through Anonymizer's Privacy Proxy Servers, masking your IP address.
3. You are now anonymous to the web site you visit and are protected from identity theft and other threats.
4. You view your requested web site anonymously and with peace of mind.

Figure 7: *Diagram of Anonymizer Anonymous Surfing [5]*

4.2.1 Why the anonymity?

According to the Tor web site, individuals need privacy for [34]:

- Privacy in web browsing – both from the remote website (so it can't track and sell your behavior), and similarly from your local ISP.
- Safety in web browsing – if your local government doesn't approve of its citizens visiting certain websites, they may monitor the sites and put readers on a list of suspicious persons.
- Circumvention of local censorship – connect to resources (news sites, instant messaging, etc) that are restricted from your ISP/school/company/government.
- Socially sensitive communication – chat rooms and web forums for rape and abuse survivors, or people with illnesses.

Using a privacy protection program protects you against a common form of Internet surveillance known as 'traffic analysis'. A basic problem for the privacy minded is that the recipient of your communications can see that you sent the communication by looking at headers. So can authorized intermediaries like Internet service providers, and sometimes unauthorized intermediaries as well [29]. A very simple form of traffic analysis might involve sitting somewhere between sender and recipient on the network, looking at headers, as demonstrated with Paros in the previous chapter. Any web administrator is able to collect information about visitors to websites. Via the IP address, it is possible to determine who the visitors are, when they visited which Internet sites, which browser and operating system they were using, and much more.

4.2.2 Anonymity tests

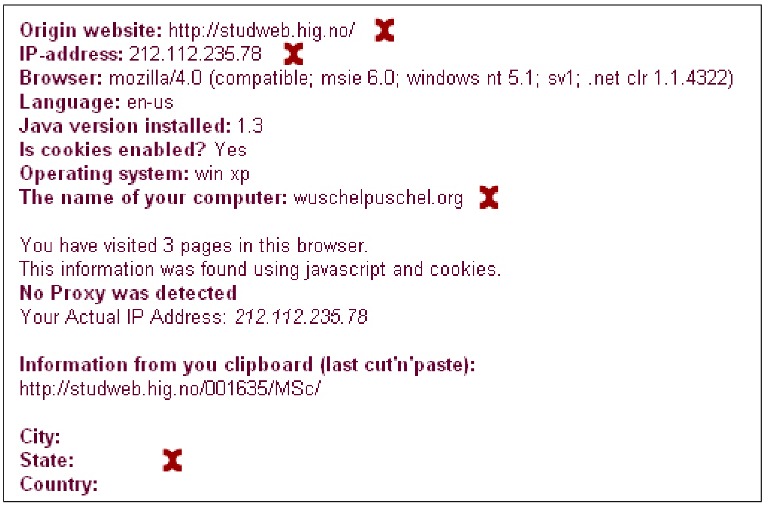
Seven different anonymisers were found, and four of these programs were chosen to be tested. These were Tor + Privoxy, Anonymizer Anonymous Surfing, ArchiCrypt Stealth and SurfAnonymous. In addition, the privacy protection feature in Norton Internet Security 2005 was tested.

Tor + Privoxy

Tor focuses only on protecting the transport of data. There is a need to use protocol-specific support software if the sites visited are not supposed to see identifying information. For example, web proxies such as Privoxy can be used while web browsing to block cookies and withhold information about the browser type, languages, etc.

Privoxy is a web proxy with advanced filtering capabilities for protecting privacy, modifying web page content, managing cookies, controlling access, and removing ads, banners and pop-ups. Privoxy has a very flexible configuration and can be customized to suit individual needs. Privoxy has applications for both stand-alone systems and multi-user networks [66].

To use Privoxy with Tor, only one option had to be changed in Privoxy. This was explained by the Tor user manual. In addition, the browser had to be set up to use Privoxy as a local proxy. Figure 8 shows the information displayed by the test web page when surfing through Tor and Privoxy. The altered information is emphasised by an *X* in the figure.



Origin website: http://studweb.hig.no/ ✗
IP-address: 212.112.235.78 ✗
Browser: mozilla/4.0 (compatible; msie 6.0; windows nt 5.1; sv1; .net clr 1.1.4322)
Language: en-us
Java version installed: 1.3
Is cookies enabled? Yes
Operating system: win xp
The name of your computer: wuschelpuschel.org ✗

You have visited 3 pages in this browser.
This information was found using javascript and cookies.
No Proxy was detected
Your Actual IP Address: *212.112.235.78*

Information from you clipboard (last cut'n'paste):
http://studweb.hig.no/001635/MSc/

City:
State: ✗
Country:

Figure 8: *Protection by Tor + Privoxy*

The Referer header (the origin web site) and the IP address were altered, as well as the name of the computer. The Referer header still displays parts of the URL, but not the details, which would be important for profiling purposes. Surfing through Tor and Privoxy, the test computer was given a range of different IP addresses, and was consequently given different geographic locations each time. The locations were among others the Netherlands, Denmark, Greece, and several states in the USA, in addition to the total suppression of geographical location as demonstrated in figure 8.

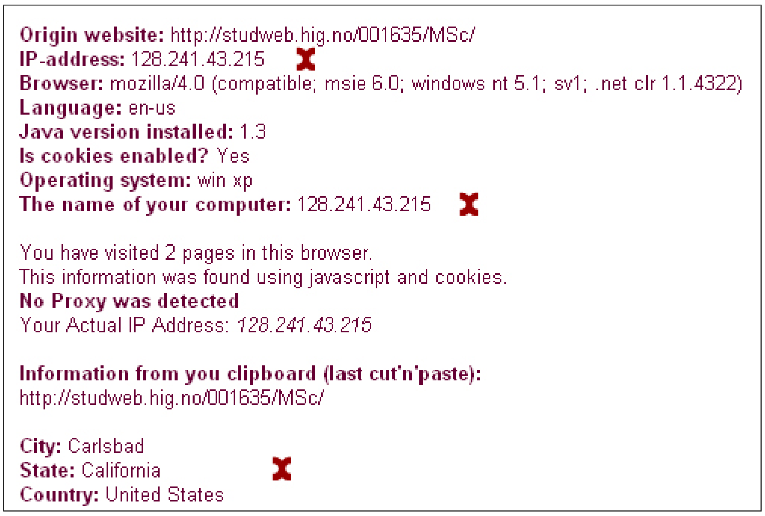
Privoxy has options to conceal the type of browser and the client operating system, as well as the Referer header, which can be blocked or forged. These services are however not default, one has to alter the configuration file to activate them. Tor and Privoxy do not have graphical user interfaces (GUIs) to help the user when setting up the software. This makes the software more difficult to use.

Anonymizer Anonymous Surfing

Anonymizer Anonymous Surfing (AAS) claims to 'protect your personal information, credit card numbers and financial transactions from online snoops' [5].

AAS works by creating an encrypted path between the computer and the Internet (as shown in Figure 7) to shield the user from online spying, phishing, and pharming. It keeps the IP address unlisted so the user can surf the Internet without being tracked, and keeps the user's online activities private. It also claims to be able to secure the data sent over a wireless connection. AAS uses 128-bit SSL technology, to ensure a high level of protection and anonymity [5]. Therefore, only the computer itself can view any data being sent to or from it.

Figure 9 shows the information displayed by the test web page when surfing through AAS. As displayed by the *X* marks in the figure, the IP address and the name of the computer were altered in this test. As a result of the IP address being altered, the geographical location is also changed correspondingly.



Origin website: http://studweb.hig.no/001635/MSc/
IP-address: 128.241.43.215 ✗
Browser: mozilla/4.0 (compatible; msie 6.0; windows nt 5.1; sv1; .net clr 1.1.4322)
Language: en-us
Java version installed: 1.3
Is cookies enabled? Yes
Operating system: win xp
The name of your computer: 128.241.43.215 ✗

You have visited 2 pages in this browser.
This information was found using javascript and cookies.
No Proxy was detected
Your Actual IP Address: *128.241.43.215*

Information from you clipboard (last cut'n'paste):
http://studweb.hig.no/001635/MSc/

City: Carlsbad
State: California ✗
Country: United States

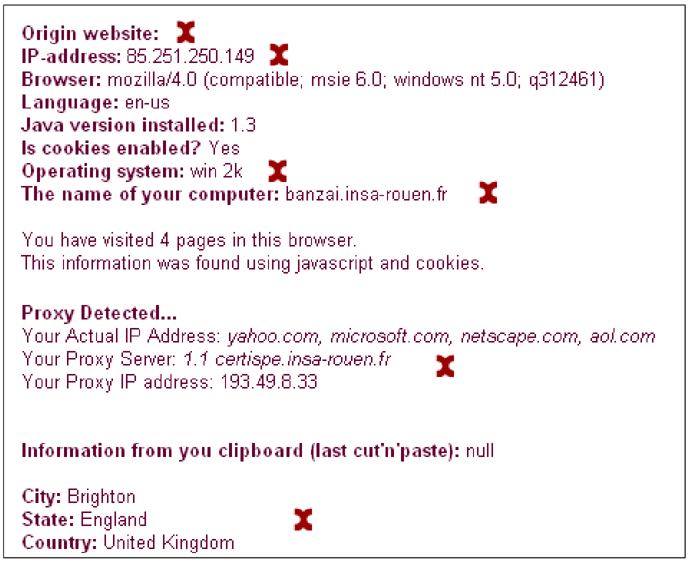
Figure 9: *Protection by Anonymizer Anonymous Surfing*

AAS was simple to install and use, and did not demand any settings being altered in the browser, as was required with Tor and Privoxy. The software worked silently in the background, seemingly without slowing down the Internet connection.

ArchiCrypt Stealth

ArchiCrypt Stealth (ACS) claims to help the users defeat local snooping, protect their privacy, and help them surf the web anonymously [80]. The software lets the user determine what data the browser should send.

Figure 10 shows the information displayed by the test web page when surfing through ACS. As usual, the IP address and the name of the computer were altered, but ACS also generated a proxy server name which was displayed. As the only program, ACS altered the operating system that was displayed, as well as completely hiding the origin web site (the Referer header). The geographical location was also altered, as a result of the IP address being changed.



Origin website: ✗
IP-address: 85.251.250.149 ✗
Browser: mozilla/4.0 (compatible; msie 6.0; windows nt 5.0; q312461)
Language: en-us
Java version installed: 1.3
Is cookies enabled? Yes
Operating system: win 2k ✗
The name of your computer: banzai.insa-rouen.fr ✗

You have visited 4 pages in this browser.
This information was found using javascript and cookies.

Proxy Detected...
Your Actual IP Address: *yahoo.com, microsoft.com, netscape.com, aol.com*
Your Proxy Server: *1.1 certispe.insa-rouen.fr* ✗
Your Proxy IP address: 193.49.8.33

Information from you clipboard (last cut'n'paste): null

City: Brighton
State: England ✗
Country: United Kingdom

Figure 10: *Protection by ArchiCrypt Stealth*

ACS claims to have the ability to alter the user's identity every second if the user wishes. By blocking or falsifying incoming and outgoing cookies, the user makes it impossible for web administrators to create a profile. The user can set which sites to accept and which ones may not display their content. Thereby, unwanted advertisers are shut out. ACS claims to filter known spyware and adware, and prevents the installation of dialers and trojans [80].

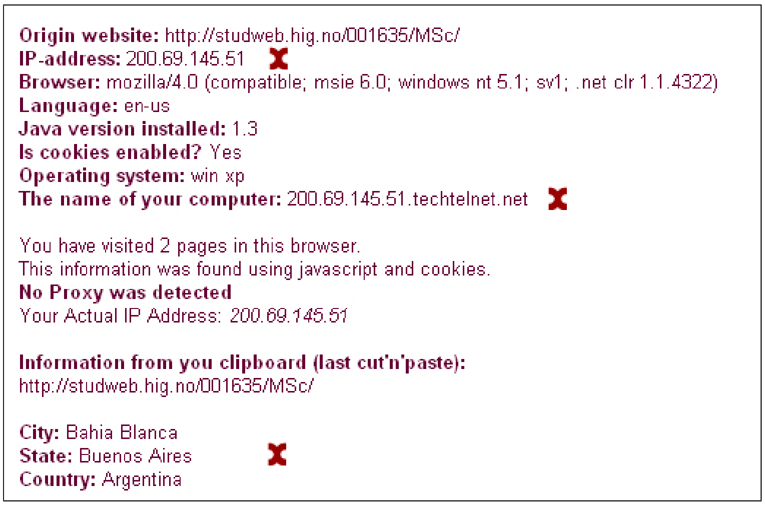
ACS offers all functions within a central and a very clear graphical user interface (GUI). No effort was needed to set up the program. It took two attempts to get any web pages through the browser, but on the second attempt it worked perfectly. This initial problem was probably due to the use of random public proxies.

SurfAnonymous

SurfAnonymous (SA) is an Internet utility that hides the IP address, thereby protecting the user from the vulnerabilities associated with this.

SA includes the following four parts: Proxy Hunter (1), Proxy Analyzer (2), Proxy Capture / IP Changer (3) and Proxy Pool (4). SA claims to be fully automated, and that the user does not need to have any knowledge of setting up proxy connections [80]. It actually demanded some more effort to set up this software. The program uses public proxies, and the user has to first collect the proxies (part 1), then export them to be validated (part 2), and after that export them once more to be used (part 4). Of course, in the full version, this is only needed the first time the program is run.

SA hides the IP address, the name of the computer and the geographic location, as shown in Figure 11.



Origin website: http://studweb.hig.no/001635/MSc/
IP-address: 200.69.145.51 ✗
Browser: mozilla/4.0 (compatible; msie 6.0; windows nt 5.1; sv1; .net clr 1.1.4322)
Language: en-us
Java version installed: 1.3
Is cookies enabled? Yes
Operating system: win xp
The name of your computer: 200.69.145.51.techtelnet.net ✗

You have visited 2 pages in this browser.
This information was found using javascript and cookies.
No Proxy was detected
Your Actual IP Address: *200.69.145.51*

Information from you clipboard (last cut'n'paste):
http://studweb.hig.no/001635/MSc/

City: Bahia Blanca
State: Buenos Aires ✗
Country: Argentina

Figure 11: *Protection by SurfAnonymous*

SA seemed to be the slowest of the programs, something which might be caused by the use of random public proxies. In Figure 11, the proxy used looks like it is South-American, but the software was tested several times, and seemed to be just as slow every time, also with other proxies.

Norton Internet Security 2005

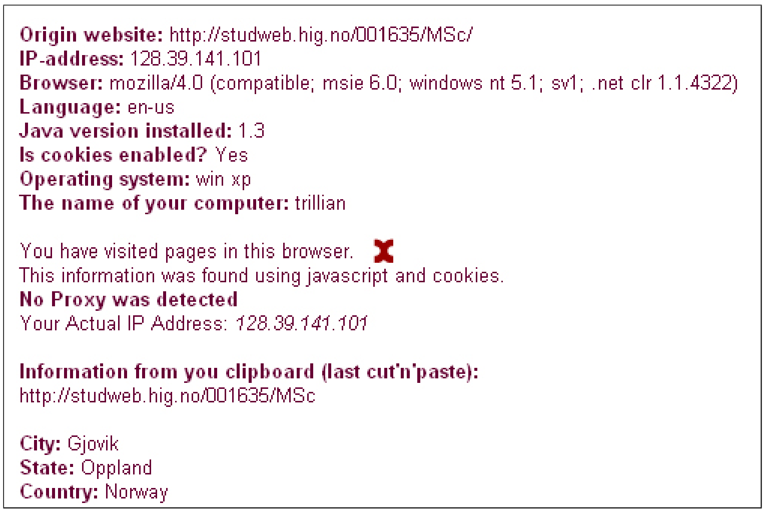
The newest version of Norton Internet Security (NIS) claims to offer a service to protect the user's privacy. The default protection level does not protect any of the information on the test web page from being revealed, but if the level of the privacy settings is changed by the user to *highest*, it allows the user to block cookies. This leads to the browser history being hidden in the test, as shown in Figure 12, as cookies were used to store this information. None of the other information was hidden when testing the web site.

Other software

Three other anonymisers were found but not tested.

E-bouncer claims to let the user boost, clean, and protect the computer, and surf the Web anonymously, as well as send secure e-mails with a hidden IP address [80]. They even supply a money-back guarantee if the program does not hide the IP address. Setting up the program demanded the same work as setting up SurfAnonymous, it just takes much longer as it searches through 11,000 public proxies, and there is no guarantee that it will find one that will protect the user's information. The trial version had several functions disabled, among others the option to make the software find suitable proxies automatically. The software was not tested, as no proxies were found when the trial version of the software was installed.

The **FindNot Internet Security Suite** claims to protect the user from being hacked or tracked, and to anonymise all web browsing (including HTTPS), email, Peer-to-Peer (P2P) file sharing, chatting (ICQ, IRC, Messenger, AIM, Yahoo etc...) and newsgroups,



Origin website: http://studweb.hig.no/001635/MSc/
IP-address: 128.39.141.101
Browser: mozilla/4.0 (compatible; msie 6.0; windows nt 5.1; sv1; .net clr 1.1.4322)
Language: en-us
Java version installed: 1.3
Is cookies enabled? Yes
Operating system: win xp
The name of your computer: trillian

You have visited pages in this browser. ✗
This information was found using javascript and cookies.
No Proxy was detected
Your Actual IP Address: *128.39.141.101*

Information from you clipboard (last cut'n'paste):
http://studweb.hig.no/001635/MSc

City: Gjovik
State: Oppland
Country: Norway

Figure 12: *Protection by Norton Internet Security*

with 128-bit encryption. It also offers an anonymous email account with 50 MB storage and anonymous file storage. The drawback is the cost, at $19.95 per month, with a minimum sign-up of one year. The service was not tested due to the limited budget of the thesis.

Complete Anonymous Web Surfing is an Internet utility to hide the user's IP address while browsing the Web. Complete Anonymous Web Surfing claims to be fully automatic, and that the user does not need to have any knowledge of setting up proxy connections [80]. Complete Anonymous Web Surfing checks the user's real IP address, verifies a number of proxy servers, deletes the non-functional proxies, sorts them by ping, selects the fastest proxy, checks the user's IP address again, and compares it with the user's real IP address. The software was not tested, as the program failed to start every time it was initiated.

4.3 Results

4.3.1 Information detection

The information detection test was originally planned with the Achilles proxy [64]. This software, however, did not seem to work with windows XP. The Paros man-in-the-middle proxy was chosen instead [16].

As the test reveals, it was not possible to see if a web page is storing information about visitors. However, it is possible to detect cookies being set, and it is possible to change the identifying information that is being sent to the web page. The only way to see if a web page is storing information about visitors, is if the company who owns the web page has posted a privacy policy saying that they do. This process of checking for privacy policies could be aided by P3P (see Chapter 2.2.1).

4.3.2 Privacy software

Figure 13 compares the tested software from the anonymity tests, with the results displaying which information was changed and which information was unaltered.

Software	PC name	IP	OS	Referer	Geo	Cookies
Tor + Privoxy	Altered	Altered	Correct	Altered	Altered	Allowed
Anonymizer	Altered	Altered	Correct	Correct	Altered	Allowed
ArchiCrypt Stealth	Altered	Altered	Altered	Altered	Altered	Allowed
SurfAnonymous	Altered	Altered	Correct	Correct	Altered	Allowed
NIS 2005	Correct	Correct	Correct	Correct	Correct	Prompted

Figure 13: *Results from the privacy test*

The software found can be divided into three groups. Norton Internet Security 2005 is a firewall with security services. Tor with Privoxy, Anonymizer Anonymous Surfing (AAS), Archicrypt Stealth, SurfAnonymous, E-Bouncer and Complete Anonymous Web Surfing are services installed on the user's computer, using proxies to replace the user's personal information. FindNot Internet Security Suite is not installed locally, and is consequently a third group, as it is a service used online.

Tor is the only free software in this test. Anonymizer Anonymous Surfing is $29.99 to buy. ArchiCrypt Stealth was free during a trial period, and is $34.75 to buy. SurfAnonymous was also free during a trial period, and is $29.95 to buy. Norton Internet Security 2005 is $49.95 to buy.

Tor is also a more advanced software, as it uses onion routing, which also protects the data. The software is not hiding the identity of the user, but protects the data the user is sending by encrypting it. The identity of the user is hidden by the use of Privoxy or similar programs. This makes Tor more reliable against attacks. AAS also uses SSL encryption to protect the information between the user and the web sites visited. Anonymizer's authorized proxies are also trusted, since in case of a deliberate information disclosure, the users know whom to blame [7].

It is also impossible to control whether or not the anonymising programs themselves are storing statistics on the users' browsing habits and other Internet use. As it does not seem possible to detect whether information is being stored by web pages, it is also not possible to detect whether information is being stored by other software. Users also have to obtain and pay for the anonymity resources somehow, except in the few cases where the service is free. This results in registration information that can also be stored in a user database.

An IP address, which is all that really identifies the web user, normally cannot be traced back to an individual, as the use of proxy servers and PPP connections means that an IP address cannot always be related to a specific machine, but to a group of computers and users [61]. In addition, an IP address might be a floating connection and can be allocated temporarily to a client's access. Location statistics based on looking up the IP address are also often misleading. This information records the country in which the user registers the IP connection, not the actual location of the user [61].

Changing the IP address could also result in negative effects. It can impact your checkbook if, for example, an e-commerce site uses price discrimination based on your country or institution of origin [34]. Some sites are also customised to suite the user, e.g. as demonstrated in `http://www.geobytes.com/`. This would not work as intended with an altered IP address. Also, when calling an emergency service using Voice over IP (Internet telephony), the localisation tool might not be working, and the help could be sent to a different location.

According to the survey performed by the European Commission in 2003 [32], only 38% of EU citizens had heard of tools or techniques limiting the collection of personal data. However, in Sweden and the Netherlands almost half of the participants had heard of such, so the knowledge varies from country to country. Around 13% had heard of *and used* such tools or techniques in Sweden, Denmark and the Netherlands. The reasons given for not using the tools among the participants who had heard of them, were that they would not know how to use them, a fear that they would be unable to install them on a computer, lack of concern about privacy issues, and a lack of conviction that the software would actually work. This thesis has shown that the programs are mostly easy to install and use, and that the software is helping the user to remain anonymous, at least to some extent. It has also shown that privacy issues should be taken into consideration.

5 Discussion

This thesis has demonstrated what kind of information it is possible to find about a visitor to a web page just by using hidden programs and scripts in the source code of the page. As it was not possible to detect whether web sites were storing information about visitors, Internet users have to rely on protection software if they wish to remain anonymous on the Internet. The software tested will help the users remain protected from profiling and data mining, by hiding and/or replacing information like the IP address, the Referer header and the name of the computer.

The goal of anonymity providing techniques is to preserve the privacy of users – who has communicated with whom, for how long, and from which location – by hiding traffic information [3]. The software tested in chapter 4.2.2 is mostly replacing the user's information with its own information, to protect the identity of the user. Some of the information, e.g. the Referer header and the user agent information, could even be replaced by the users themselves by the use of a man-in-the-middle proxy. However, by hiding the user agent information, one could actually create difficulties displaying some web pages, as these are often adapted to suit the different browsers, like Microsoft Internet Explorer, Mozilla Firefox, Netscape, Safari, Konqueror and Opera.

5.1 Vulnerabilities to anonymising services

The software that was tested protects the user from revealing unnecessary information. However, the software also has weaknesses. For example, the range of nodes being a part of each service is often limited. This might lead to the anonymising service being revealed by the IP address in question. Accordingly this traffic is not worth noticing when creating statistics for profiling. Some services have even started denying access when using an anonymising network, by enumerating the nodes in question. Services like Slashdot (`http://slashdot.org/`) and Wikipedia (`http://www.wikipedia.org/`) could have huge challenges trying to prevent anonymous postings to their newsgroups, and would therefore benefit if anonymising services were not allowed.

There is a need to avoid that the limited number of nodes is leading to the disclosure of the use of anonymisers. One could ask if it is possible to hide the IP address and other personal information completely instead of just replacing it with someone else's information. However, this might again lead to the denial of access to certain services.

5.1.1 Attack resistance

Although various MIX-based anonymity systems have been operational, none of them delivered many quantitative results about performance, bandwidth overhead, or other issues that arise when implementing or operating such a system [67]. So far, almost every commercial privacy technology venture has failed, like e.g. PipeNet and Rewebber, with Anonymizer being a notable exception [45]. Compared to other infrastructure-heavy attempts, Anonymizer has a relatively simple architecture, at the expense of protecting against a weaker threat model. However, it seems that this weaker threat model is sufficient for most consumers [45].

Argyrakis et al. [7] have analysed nine different security threats to anonymising services, and the different vulnerabilities. They also conclude that Onion Routing, which Tor is based on, gives the better protection. The proxy-based services offer less protection as they are vulnerable to several attacks, like the statistical logging attacks described by Sun et al. [74].

The MIX based approaches all use combinations of some form of encryption and interposition of intermediaries, and impose a rigid structure on the *shape* of traffic (synchronised transfer of standard-length blocks of data, and even dummy 'covering traffic') to prevent leakage of information through timing or size of messages [74]. Onion Routing dispenses with measures such as synchronisation and covering traffic, but is thus also vulnerable to the kind of traffic analysis described in Sun et al. [74]. This is the result of a tradeoff between efficiency in terms of bandwidth, latency and server load, and effectiveness at disguising traffic.

Tor addresses limitations in the original Onion Routing design by adding perfect forward secrecy, congestion control, directory servers, integrity checking, configurable exit policies, and a practical design for location-hidden services via rendezvous points [29]. Like all anonymising networks that are fast enough for web browsing, Tor does not provide protection against end-to-end timing attacks: If your attackers can watch the traffic coming out of your computer, and also the traffic arriving at your chosen destination, they can use statistical analysis to discover that they are part of the same circuit [29].

The web-oriented approaches tend to downplay or eliminate the role of encryption, assuming an adversary with limited traffic observation powers, and relying instead on techniques such as trusted (or multiple random) intermediaries, pseudonyms and multicast to disguise the identities of browsing users [74]. Some do not use encryption at all, others use encryption between browser and proxy to foil eavesdroppers but do not attempt end-to-end encryption.

5.2 Is profiling positive use?

Profiling techniques allows marketers to better target customers who may be amenable to certain offers, or they may help marketers create products which will suite the consumers better. Personalised purchase recommendations on a web site can significantly increase the likelihood of a customer making a purchase, compared to unpersonalised suggestions [13]. Most online vendors collect buying information about their customers, and make reasonable efforts to keep this data private. However, customer data is a valuable asset, and it is routinely sold as such, for example when companies have suffered bankruptcy [13].

When the online retailer ToySmart went bankrupt in 2000, the bankruptcy receiver put the ToySmart database of customer information up for sale. However, ToySmart had promised in its privacy policy to never give its customer data to any third party, ever. The US Federal Trade Commission had a mandate to make sure that ToySmart honoured the promises made in its privacy policy, while the bankruptcy court had a mandate to sell as many assets as it could to cover debts. In the end, Disney agreed to buy some of the assets, including the customer list, which Disney promised to promptly destroy as a condition of the sale [57, p. 20].

Choicepoint claims to be USA's leading provider of identification and credential verification services [17]. In fact, the company sold the personal information of nearly 145,000 people to inadequately vetted bogus businesses [71]. The people whose information was compromised were not customers of ChoicePoint, just accidental citizens of the vast databases of the Georgia-based information broker. The way that ChoicePoint behaved after the breach did not encourage trust: From an initial, bumbling response that smacked of marketing, to a changing story about what had happenedand how the company was responding.

> crimes **committed against** ChoicePoint that may have resulted in personally identifiable information such as name, address, Social Security number or credit report being viewed by businesses that should not have accessed such information [17]

The incident included one last twisted bit of irony: ChoicePoint chairman and CEO Derek V. Smith had recently written two books about how individuals can protect themselves in the information age [71].

The 'dossier effect' is dangerous – when it is so easy to build a comprehensive profile of individuals, many will be tempted to take advantage of it, whether for financial gain, vicarious entertainment, illegitimate purposes or other unauthorised use [46]. It might be positive for the user to receive recommendations when logging on to a web site, but in the long term, it might be better if no profiles were made. If the users want the books, music or other goods that are recommended, they will probably be able to find them anyway.

5.2.1 Adversaries

Outside of the Internet, anonymity is widely accepted and recognised as valuable in today's society [39]. On the other side, anonymity can also be misused to commit criminal activities without leaving any traces. As *the New Yorker* magazine, cited in Goldberg et al. [46], explained in a cartoon: 'On the Internet, nobody knows you're a dog'. Illicit use of anonymity is all too common on the Internet. Like most technologies, Internet anonymity techniques can be used for better or worse, so it should not be surprising to find some unfavourable uses of anonymity. For instance, anonymity tools are used to distribute copyrighted software without permission, and e-mail spammers are learning to take advantage of anonymity techniques to distribute their marketing ploys widely without retribution [46].

Widespread availability of anonymity will mean that site administrators will have to rely more on first-line defences and direct security measures rather than on the deterrent of tracing [46]. With systems that are providing anonymity, it is no longer possible to track certain computer crimes to the person responsible. However, the ban of privacy technologies, or restrictions to use them, is not the right solution for this problem, because it would severely restrict the users' possibilities to protect their privacy. Criminals, however, would still find other means or could set up their own technologies to commit crimes without leaving relevant traces [39].

Knowing the source and destination of your Internet traffic allows others to track your behaviour and interests [29]. It can even threaten your job and physical safety by revealing who and where you are. For example, if you're travelling abroad and you connect to your employer's computers to check or send mail, you can inadvertently reveal your national origin and professional affiliation to anyone observing the network, even if the connection is encrypted [34]. Also, there might be an increasing need for anonymity services as the military forces and government agencies are using an increasing amount of computer equipment. Traffic flow confidentiality is always important in these cases, as it should be impossible to trace who is talking to whom. In addition, agents who are abroad on commissions should be able to 'phone home' without being disclosed, both when using normal telephones and when using Voice over IP.

5.3 Recommendations

For the 'normal' Internet user, the utilitarian value of anonymising programs is limited. As many ISP's use dynamic IP addresses, there is normally no need to hide the IP address. However, as more and more users switch to broadband connections with routers, maintaining a lasting Internet connection, it might become more interesting to have an option to hide from profiling and statistical information being collected. Still, most ISPs keeps logs on which user was connected through which IP address at each time. These logs are often kept for a longer period of time than expected (and/or allowed), as logs are not easy to wipe out. With IP addresses which are no longer dynamic, the users will probably become more aware of the traceability. However, the dynamic addresses are just as traceable. The difference is in the availability and traceability for third-party-services like Google, eBay and Messenger.

Anonymisers are useful tools to ensure that identifying information is not transferred during online interactions in which no personal information need to be revealed. However, they are of limited use for transactions in which personal information must be explicitly revealed [21]. Most online purchases are for example made with credit cards, which identify the individual and facilitate the collection of purchasing data. The lack of a cash equivalent in the online world, and its reduced use in the physical world, will seriously alter the privacy of individuals' financial dealings [9]. As long as the transfer of personal data remains a requirement for most routine transactions, privacy enhancing technologies will offer only limited privacy protection, except technologies which also use data encryption, like Tor and Anonymizer Anonymous Surfing. Technology could be used to a much greater privacy enhancing benefit if the architecture of our transaction systems were changed to support transactions that reveal much less information. While such changes are technically feasible, there appear to be significant obstacles to adoption [21].

In some cases, protection against third-party cookies could be sufficient for a user. This can be achieved just by changing the settings in the user's browser or firewall software. This would prevent profiling by companies such as DoubleClick.

6 Conclusion

This thesis started out with two questions:

- How much information is it possible to find about a visitor to a web page just by using hidden programs and scripts in the source code of the page?
- Is it possible to detect and/or protect oneself against this information being revealed?

The work has shown that it is quite easy to find information about visitors to a web page. It has also been demonstrated how it is possible to protect oneself from revealing this information by using proxies and/or anonymisers.

The information that was found was the Referer header, the IP address, browser name and version, Java version installed, the operating system, part of the browser history, the name of the computer, the clipboard content and the geographical location. To extract the browser information, the operating system and the IP address was quite easy, as most web scripting manuals have this information. To find the clipboard content, the geographical location and the browser history it was necessary to do a bit more research. This shows that to be able to extract this last information mentioned, one needs to be really determined to find it. Accordingly, it demands more knowledge, but as the search engines are improved, it will also be easier to use these to find such information.

The experiment found that it was not possible to detect whether or not a web page is storing information about its visitors. By the use of a man-in-the-middle proxy it was however possible to detect which cookies are being set, and it is possible to alter information in the HTTP header. This could also be accomplished in some browsers, as the browser could be set to warn and/or prompt the user of every cookie that is being set on the browser. The P3P technology could also help aiding the user in cases where the web site has a P3P-encoded privacy policy. However, this technology does not guarantee that the web site owners do as they promise, only that they have made an effort to write a policy. Still, to have P3P-encoded the privacy policies shows that the companies in question are interested in making a good impression towards the public.

When protecting a user, the anonymising software removes the user's information (IP address, name of computer, etc.), and replaces it with its own information or the information from one of its nodes/proxies. Two of the anonymisers that were tested also altered the Referer header, Tor + Privoxy and ArchiCrypt Stealth. However, most of the programs in the test will only protect the identity of the user, not the data that is being sent. Only Tor and Anonymizer Anonymous Surfing use cryptographic protection of the data as well. These would accordingly also be of use when transferring personal data for use in online transactions where the user is not already protected by HTTPS.

Someone once pointed out to the head of Sun Microsystems, Scott McNealy, that a new Internet technology they were developing could lead to a massive invasion of people's privacy [47]. His response was supposed to have been: 'You have zero privacy now. Get over it.’ It is true that privacy and security can be endangered when using the Internet. However, it is still possible to fight back, and at least make some improvements!

Ultimately, the ideal and secure computer is a system without any contact to the outside world. That is, no disks or CD-drives, and no direct contact to other computers by modem or network adapter. [72]

As Internet users become aware of what information is revealed, they hopefully also see the possibility of protection. This thesis has demonstrated some of the methods that can be used by ‘ordinary’ Internet users. Not using the Internet is not an option.

6.1 Future work

During the work with the thesis, several new questions surfaced.

One of the questions is how it is possible for the user to con the server. This thesis has concentrated on the cases where the server exploits visitors’ information. One could say that when using an anonymiser to change user information, the user is tricking the server into believing that the user is someone else, but which other possibilities exist? An example is described in Kabir [54, p. 27], where the input in an address field is changed to make the output in the browser become the password file from the server (by adding cat /etc/passwd in the address line). It would be important to know if these possibilities still can be used, and if there are more possibilities like this.

The problem of fighting phishing and pharming has also become very interesting. Ollmann [63] answers some of the questions regarding phishing, but there are still more questions to look into, for example if it is possible to automate the protection.

Using proxies to hide an Internet user’s identity has been said to be vulnerable to several attack. There are several reports on different attacks on anonymising services / privacy enhancing technologies, but they all seem to concentrate on the more advanced technologies, saying that the simpler, commercially available, services are vulnerable. It ought to be tested if it is possible to find information *behind* the proxy server.

As it seems like software does not exist which is able to detect information being stored, there still is a need to find out if it is possible to make this kind of software. It was not possible to find out how to make such a program during this thesis work, but it might still be possible. What kind of variables could be detected? What would it take to detect whether information is stored?

One method used to protect against surfing habits and personal information being revealed is the dynamic URL. The question remains how long the URL must be to supply any protection, and how many bits of the URL must be dynamic. Another advantage with the dynamic URL is the time-stamp function, which can log the user out from a web site due to inactivity.

Bibliography

[1] Aadland, C. [2005], 'Usynlige spor (invisible tracks)', *Teknisk Ukeblad* **152**(14), 16–21.

[2] Abrams, M. [2003], Choice, *in* P. J. Bruening, ed., 'Considering consumer privacy: A resource for policymakers and practitioners'. http://www.cdt.org/privacy/ccp/.

[3] Agrawal, D., Kesdogan, D. and Penz, S. [2003], Probabilistic treatment of mixes to hamper traffic analysis, *in* B. Werner, ed., 'Proceedings of the 2003 IEEE Symposium on Security and Privacy', The Printing House.

[4] Allen, J. and Hornberger, C. [2002], *Mastering* PHP *4.1*, Sybex, Alameda.

[5] Anonymizer [2004], 'Anonymizer website'. http://www.anonymizer.com, April 22 2005.

[6] Antón, A. I. and Earp, J. B. [2004], 'A requirements taxonomy for reducing web site privacy vulnerabilities', *Requirements Engineering* **9**(3), 169–185.

[7] Argyrakis, J., Gritzalis, S. and Kioulafas, C. [2004], 'Privacy enhancing technologies: A review', *Lecture Notes in Computer Science* **2739/2004**, 282–287.

[8] Arnold, S. E. [2002], Internet users at risk: The identity/privacy target zone, *in* A. P. Mintz, ed., 'Web of Deception: Misinformation on the Internet', CyberAge Books, New Jersey.

[9] Berman, J. and Mulligan, D. [1999], 'Privacy in the digital age: Work in progress', *Nova Law Review* **23**(2). http://www.cdt.org/publications/lawreview/1999nova.shtml, May 4 2005.

[10] Booth, W. C., Colomb, G. G. and Williams, J. M. [2003], *The craft of research*, 2nd edn, University of Chicago Press, Chicago.

[11] Broder, A. J. [2003], 'Data mining, the internet, and privacy', *Lecture Notes in Computer Science* **1836/2000**, 56–73.

[12] Bygrave, L. A. [2002], *Data Protection Law: Approaching its rationale, logic and limits*, Kluwer Law International, The Hague.

[13] Canny, J. [2002], Collaborative filtering with privacy, *in* A. D. Williams, ed., 'Proceedings of the 2002 IEEE Symposium on Security and Privacy', The Printing House.

[14] CDT [2004], 'CDT's guide to online privacy'. http://www.cdt.org/privacy/guide/introduction/, May 3 2005.

[15] Chesbro, M. [2000], *The complete guide to* E*-security*, Paladin Press, Colorado.

[16] Chinotec [2004], 'Paros - for web application security assessment'. `http://www.parosproxy.org/index.shtml`, April 4 2005.

[17] ChoicePoint [2005], 'ChoicePoint – identification and credential verification'. `http://www.choicepoint.com/`, June 1 2005.

[18] Chuck, L. B. [2002], Welcome to the dark side: How e-commerce, online consumer, and e-mail fraud rely on misdirection and misinformation, *in* A. P. Mintz, ed., 'Web of Deception: Misinformation on the Internet', CyberAge Books, New Jersey.

[19] Clayton, G. and Stewart, A. [2003], The privacy management process, *in* P. J. Bruening, ed., 'Considering consumer privacy: A resource for policymakers and practitioners'. `http://www.cdt.org/privacy/ccp/`.

[20] Cosentino, C. [2001], *Essential* PHP *for web professionals*, Prentice Hall, New Jersey.

[21] Cranor, L. F. [2003], The role of privacy enhancing technologies, *in* P. J. Bruening, ed., 'Considering consumer privacy: A resource for policymakers and practitioners'. `http://www.cdt.org/privacy/ccp/`.

[22] Creswell, J. W. [2003], *Research design: Qualitative, quantitative, and mixed method approaches*, 2nd edn, Sage Publications, London.

[23] Culnan, M. J. [2003], How privacy notices promote informed consumer choice, *in* P. J. Bruening, ed., 'Considering consumer privacy: A resource for policymakers and practitioners'. `http://www.cdt.org/privacy/ccp/`.

[24] Danezis, G. [2004], Better Anonymous Communications, PhD thesis, University of Cambridge. `http://www.cl.cam.ac.uk/~gd216/thesis.pdf`, June 17 2005.

[25] Demuth, T. [2002], A passive attack on the privacy of web users using standard log information, *in* '2002 Workshop on Privacy Enhancing Technologies', San Francisco.

[26] Demuth, T. and Rieke, A. [1998], 'Anonym im world wide web?', *Datenschutz und Datensicherheit (DuD)* **11**, 623–627.

[27] Demuth, T. and Rieke, A. [2000], Bilateral anonymity and prevention of abusing logged web addresses, *in* '2000 Military Communications International Symposium', Los Angeles.

[28] DeVault, J., Tretick, B. and Ogorzelec, K. [2003], Privacy and independent verification: What consumers want, *in* P. J. Bruening, ed., 'Considering consumer privacy: A resource for policymakers and practitioners'. `http://www.cdt.org/privacy/ccp/`.

[29] Dingledine, R., Mathewson, N. and Syverson, P. [2004], Tor: The second-generation onion router, *in* 'Proceedings of the 13th USENIX Security Symposium'. `http://tor.freehaven.net/tor-design.pdf`, June 17 2005.

[30] DNSSEC [2005], ‘Dns security extensions: Securing the domain name system’. http://www.dnssec.net/, June 17 2005.

[31] EC [1997], ‘Information technology and data privacy’. http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_109_en.pdf, Jan. 1997.

[32] EC [2003], ‘Data protection’. http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_196_data_protection.pdf, Dec. 2003.

[33] EC [2004], ‘Issues relating to business and consumer e-commerce’. http://europa.eu.int/comm/public_opinion/archives/ebs/ebs_201_executive_summary.pdf, March 2004.

[34] EFF [2005], ‘Tor’. http://tor.eff.org/, April 22 2005.

[35] EPIC [1997], ‘Surfer beware: personal privacy and the internet’. http://www.epic.org/reports/surfer-beware.html, June 1997.

[36] EPIC [1998], ‘Surfer beware II: notice is not enough’. http://www.epic.org/reports/surfer-beware2.html, June 1998.

[37] EPIC [1999], ‘Surfer beware III: privacy policies without privacy protection’. http://www.epic.org/reports/surfer-beware3.html, Dec. 1999.

[38] FindNot [2004], ‘Surfing anonymous’. http://www.findnot.com/?1_surfing_anonymous, April 21 2005.

[39] Fischer-Hübner, S. [2003], ‘Privacy in the global information society’, *Lecture Notes in Computer Science* **1958/2001**, 5–33, 107–165.

[40] Futsæter, K.-A. [2001], ‘State of the internet’. http://www.tns-gallup.no/, Oct. 26 2000.

[41] GeoDirection [2004], ‘GeoDirection from GeoBytes’. http://www.geobutton.com/GeoDirection.htm, Feb. 22 2005.

[42] Germaise, S. [2003], *Privacy tactics*, Tetra Mesa, New York.

[43] GetNetWise [2004], ‘GetNetWise - privacy’. http://privacy.getnetwise.org/, April 20 2005.

[44] Gilmore, W. J. [2001], *A programmer’s introduction to* PHP *4.0*, Apress, Berkeley.

[45] Goldberg, I. [2002], Privacy-enhancing technologies for the internet, ii: Five years later, *in* R. Dingledine and P. Syverson, eds, ‘Proceedings of Privacy Enhancing Technologies workshop (PET 2002)’, Springer-Verlag, LNCS 2482. http://freehaven.net/anonbib/papers/petfive.pdf.

[46] Goldberg, I., Wagner, D. and Brewer, E. [1997], Privacy-enhancing technologies for the internet, *in* ‘Proceedings of the 42nd IEEE Spring COMPCON’, IEEE Computer Society Press.

[47] Gralla, P. [2002], *The Complete Idiot's Guide to Internet Privacy and Security*, Pearson Education, Indianapolis.

[48] Ham, S. [2003], Internet privacy: The case for pre-emption, *in* P. J. Bruening, ed., 'Considering consumer privacy: A resource for policymakers and practitioners'. `http://www.cdt.org/privacy/ccp/`.

[49] Horstmann, C. S. and Cornell, G. [2001], *Core* JAVA *2: Volume* I *Fundamentals*, Sun Microsystems Press, Palo Alto.

[50] Hughes, S. [2001], *PHP Developer's Cookbook*, Sams, Indianapolis.

[51] Huseby, S. H. [2004], *Innocent Code*, John Wiley & Sons, Chichester.

[52] ITS [1996], 'Telecommunications: Glossary of telecommunication terms', Federal standard 1037 C. `http://www.atis.org/tg2k/`, March 12 2005.

[53] JanetSystems [2005], 'Browser history'. `http://www.janetsystems.co.uk/Default.aspx?tabid=72&itemid=7`, Feb. 20 2005.

[54] Kabir, M. J. [2003], *Secure* PHP *Development*, Wiley Publishing, Indianapolis.

[55] Lawler, B. and Cooper, S. [2003], The opt-in approach to choice, *in* P. J. Bruening, ed., 'Considering consumer privacy: A resource for policymakers and practitioners'. `http://www.cdt.org/privacy/ccp/`.

[56] Lemmey, T., Klein, S. and Neumann, T. [1999], 'Architecture is policy'. `http://www.eff.org/Privacy/?f=19990406_privacypaper.html`, April 1999.

[57] Levine, J. R., Everett-Church, R. and Stebben, G. [2002], *Internet Privacy for Dummies*, Wiley Publishing, New York.

[58] Miller, M. [2002], *Absolute PC Security and Privacy*, Sybex, Alameda.

[59] Murphy, J., Hofacker, C. F. and Bennett, M. [2001], 'Website-generated market-research data: Tracing the tracks left behind by visitors', *Cornell Hotel and Restaurant Administration Quarterly* **42**, 82–91.

[60] NAI [2001], 'Network advertising initiative'. `http://www.networkadvertising.org/`, May 20 2005.

[61] Nicholas, D. and Huntington, P. [2003], 'Micro-mining and segmented log file analysis: A method for enriching the data yield from internet log files', *Journal of Information Science* **29**, 391–404.

[62] NRC [2000], 'Electronic information exchange glossary'. `http://www.nrc.gov/site-help/eie/terms_id.html`, March 12 2005.

[63] Ollmann, G. [2004], 'The phishing guide: Understanding & preventing phishing attacks'. `http://www.ngssoftware.com/papers/NISR-WP-Phishing.pdf`, Sep. 2004.

[64] Packetstorm [2004], 'Achilles web proxy'.
http://packetstorm.linuxsecurity.com/web/, April 4 2005.

[65] PandaSoftware [2005], 'Panda software: Glossary'.
http://www.pandasoftware.com/virus_info/glossary/, June 5 2005.

[66] Privoxy [2004], 'Privoxy privacy protecting web proxy'.
http://www.privoxy.org/, April 22 2005.

[67] Rennhard, M., Rafaeli, S., Mathy, L., Plattner, B. and Hutchison, D. [2002], Analysis of an anonymity network for web browsing, *in* 'Proceedings of the IEEE 7th Intl. Workshop on Enterprise Security (WET ICE 2002)', Pittsburgh, pp. 49–54.
http://www.tik.ee.ethz.ch/~rennhard/publications/WetIce2002.pdf.

[68] Rodriguez, S. [2003], 'Clipboard exploit'.
http://www.arstdesign.com/articles/clipboardexploit.html, Feb. 23 2005.

[69] Rust, R. T., Kannan, P. K. and Peng, N. [2002], 'The customer economics of internet privacy', *Journal of the Academy of Marketing Science* **30**, 455–464.

[70] Salkind, N. J. [2003], *Exploring Research*, 5th edn, Pearson higher education, Essex.

[71] Scalet, S. D. [2003], 'CSOonline: The five most shocking things about the choicepoint debacle'.
http://www.csoonline.com/read/050105/choicepoint.html, May 2005.

[72] Schaeffer, F. [2002], *Surfing Anonymously*, Data Becker, Newton, MA.

[73] Sultan, F. [2002], 'Consumer response to the internet: an exploratory tracking study of on-line home users', *Journal of Business Research. Elsevier Science* **55**, 655–663.

[74] Sun, Q., Simon, D. R., Wang, Y.-M., Russell, W., Padmanabhan, V. N. and Qiu, L. [2002], Statistical identification of encrypted web browsing traffic, *in* A. D. Williams, ed., 'Proceedings of the 2002 IEEE Symposium on Security and Privacy', The Printing House.

[75] Ullman, L. [2002], PHP *advanced for the world wide web*, Peachpit Press, Berkeley.

[76] Unanue-Zahl, P. [2003], 'Kredittkortsvindel florerer på nettet (credit card fraud flourishing on the internet)'. http://www.vg.no/pub/vgart.hbs?artid=202380, Nov. 20 2003.

[77] W3C [2004], 'Platform for privacy preferences (P3P) project'.
http://www.w3.org/P3P/, May 11 2005.

[78] W3Schools [2004], 'W3Schools online web tutorials'.
http://www.w3schools.com, Feb. 21 2005.

[79] WebEraser [2004], 'Web Eraser'. http://www.weberaser.com/, April 20 2005.

[80] ZDNet [2005], 'ZDNet - where technology means business'.
http://downloads-zdnet.com.com/, April 26 2005.

Appendices

A Terminology

- Affinity marketing: Once a person is placed in a cluster, mathematical algorithms can predict certain patterns or predispositions of behaviour for the group. No individual action can be predicted, but in an affinity group, a certain number of individuals will adopt the predicted behaviour. Affinity group marketing, therefore, allows a person in a group who bought X to be offered product Y. The marketer knows a certain number of people will buy Y because they bought X [8].
- Agents: Scripts that perform specific tasks and are equipped with some type of mechanism that allows the script to take different actions depending upon a situation [8].
- ASCII text: *American Standard Code for Information Interchange* – for representing characters (letters, numbers, punctuation marks, etc.) as numbers [65].
- CoE Convention: The CoE *Convention for the Protection of Individuals with regard to Automatic Processing of Personal Data*.
- Client: In networking, a software application that allows the user to access a service from a server computer, e.g., a server computer on the Internet [52].
- Cookie: A handle, transaction ID or other token of agreement between cooperating programs [8]. Bits of data put on the computer by a web server, that can be used to track surfing habits.
- Data: Programs, files, and other information stored in, communicated, or processed by a computer [62].
- Data mining: A series of routines that look at data, make decisions about how the data relate, and then outputs reports driven by the content of large collections information, collections too large for individuals to review as productively – for example, a year's collection of American Express credit card users' transactions [8].
- Database: A set of related information created, stored, or manipulated by a computerized management information system [62].
- Directive 2002/58/EC: The EC Directive *on Privacy and Electronic Communications*.
- Directive 94/46/EC: The EC Directive *on the Protection of Individuals with Regard to the Processing of Personal Data and on the Free Movement of Such Data*.
- DNS (Domain Name System): System to enable communication between computers connected across a network or the Internet. It means that computers can be located and assigns comprehensible names to their IP addresses [65].
- EU: The European Union.
- HTML (Hyper Text Markup Language): The language used on the World Wide Web. Web browsers decode HTML and display the page [52].

- HTTP (Hyper Text Transfer Protocol: A communication system that allows web pages to be viewed through a browser [65].
- IETF (Internet Engineering Task Force): An all volunteer organization responsible for publishing RFCs and Internet Standards [52].
- Information: Here used in the same meaning as data (see this).
- IP address: The numerical address of something on the Internet, the address that computers understand. A series of four numbers, separated by periods, for example *128.39.141.138* [47].
- ISP (Internet Service Provider): A company that offers access to the Internet and other related services [65].
- OECD: Organization for Economic Cooperation and Development.
- OECD Guidelines: The OECD *Guidelines Covering the Protection of Privacy and Transborder Flows of Personal Data*.
- Opt-in marketing: With permission to resell or use the address for direct marketing of other products and services [8].
- Packet sniffing: Monitoring network traffic in order to recognize and decode certain packets of interest [52]).
- PC: Personal Computer.
- Personal data: Article 2(a) of the CoE Convention defines *personal data* as *any information relating to an identified or identifiable individual*. Exactly the same definition is given in paragraph 1(b) of the OECD Guidelines [12]. Article 2(a) of the EC Directive defines *personal data* as:

  > any information relating to an identified or identifiable natural person; an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural or social identity.
- PPP connection: Point-to-point connection. A configuration where there are only two access points [52].
- Privacy: The ability of individuals to exercise control over the disclosure and subsequent uses of their personal information (Westin 1967, quoted in Culnan [23]).
- RFC (Request for Comment): The document series, begun in 1969, which describes the Internet suite of protocols and related experiments. Not all (in fact very few) RFCs describe Internet standards, but all Internet standards are written up as RFCs [52].
- Server: A computer system that responds to requests from client systems [62].
- Sniffer: A script that looks for words, phrases, terms, concepts, and tendencies in digital messages. Separate software is required to interpret what the sniffer senses [8].
- Spoofing: Making a message or process appear to come from another data source. Because systems and users ‘trust’ known sources, spoofing allows a wrongdoer to enter the target system [8].

- SSL (Secure Socket Layer): a method in which all information between the user and a website is encrypted [5].
- UN Guidelines: The United Nations' *Guidelines Concerning Computerized Personal Data Files*.
- URL (Uniformed Resource Locator): A standardized device for identifying and locating certain records and other resources located on the World Wide Web [62].
- US, the: The United States of America, USA.
- User: A person, organization, or other entity (including a computer or computer system), that employs the services provided by a telecommunication system, or by an information processing system, for transfer of information [52].

B Test web page

E-mail sent to participants:

Heisann!!

Nå sitter jeg altså og leker med master-oppgave, og da trenger jeg litt hjelp... Jeg har laget en side der jeg viser litt om hvor enkelt en kan finne informasjon om besøkende til en webside. Denne siden finnes på: http://studweb.hig.no/001635/MSc/

For at jeg skal få brukbar statistikk, er det litt greit at jeg får noen besøkende på siden, og der kommer dere inn i bildet! Det eneste dere trenger å gjøre er å trykke på linken (evt kopiere den over i nettleser for de som foretrekker det), lese litt info, og trykke på to knapper...

Og til slutt en forsikring: Jeg lagrer ikke noen personlig informasjon, mine statistikker registrerer OM jeg får tak i informasjonen, ikke informasjonen i seg selv.

På forhånd tusen takk for hjelpen!

Translation:

Hi!!

I am now currently working on my MSc thesis, and now I need some assistance... I have made a web page where I demonstrate how easy it is to find information about visitors to a web page. This page can be found at: http://studweb.hig.no/001635/MSc/

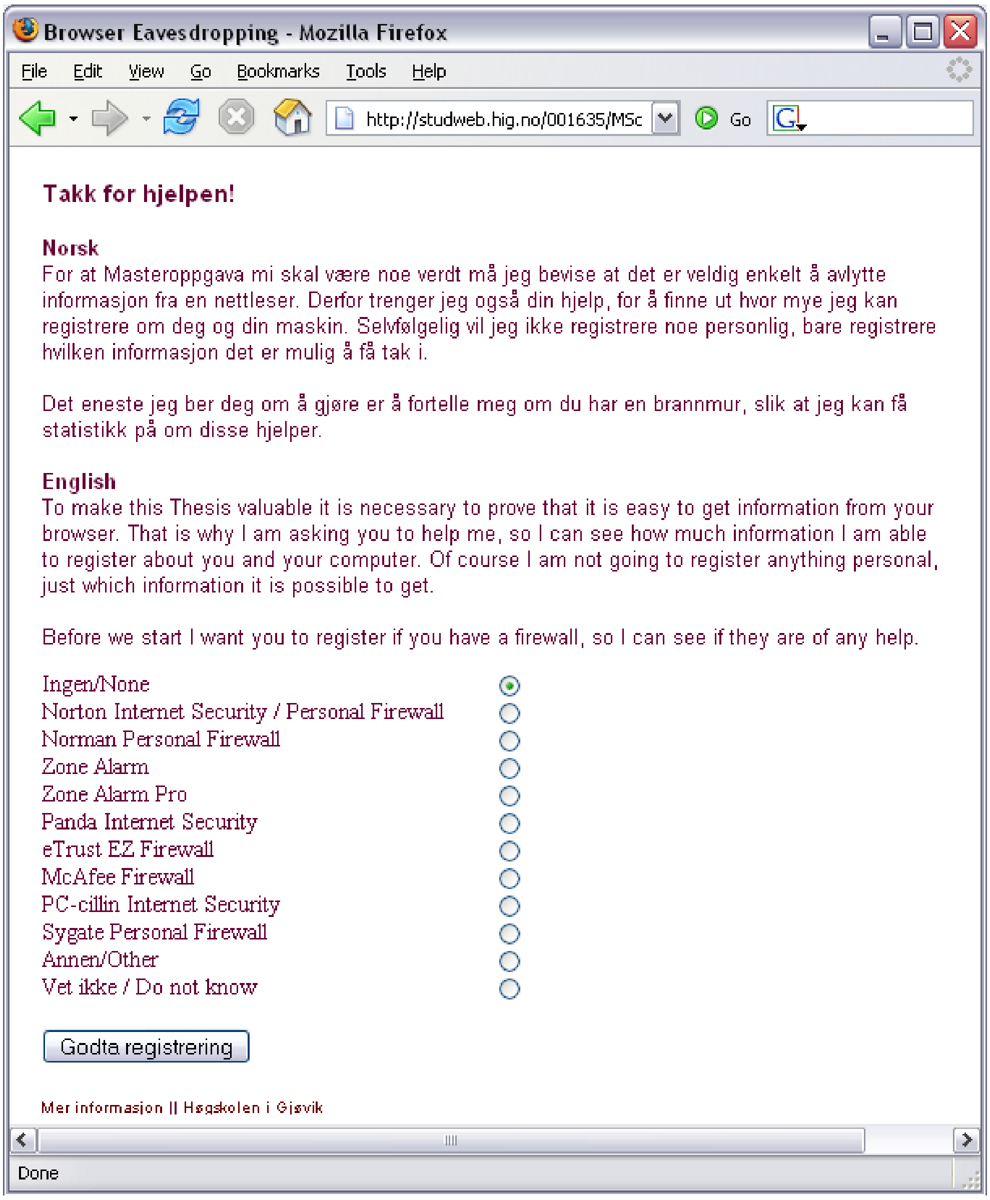
To achieve a usable statistics, it would be an advantage to have a few visitors to this web page, and this is where I want you to help. The only thing you need to do is to press the link (or copy it to your browser if you prefer this methos), read some information, and press two buttons...

And finally an assurance: I am not storing any personal information, my statistics will only register IF I am able to get the information, not the information itself.

Thank you very much in advance!

Web page:

Figure 14 shows the first page as displayed when a participant entered.


Browser Eavesdropping - Mozilla Firefox
File Edit View Go Bookmarks Tools Help
http://studweb.hig.no/001635/MSc
Go

Takk for hjelpen!

Norsk
For at Masteroppgava mi skal være noe verdt må jeg bevise at det er veldig enkelt å avlytte informasjon fra en nettleser. Derfor trenger jeg også din hjelp, for å finne ut hvor mye jeg kan registrere om deg og din maskin. Selvfølgelig vil jeg ikke registrere noe personlig, bare registrere hvilken informasjon det er mulig å få tak i.

Det eneste jeg ber deg om å gjøre er å fortelle meg om du har en brannmur, slik at jeg kan få statistikk på om disse hjelper.

English
To make this Thesis valuable it is necessary to prove that it is easy to get information from your browser. That is why I am asking you to help me, so I can see how much information I am able to register about you and your computer. Of course I am not going to register anything personal, just which information it is possible to get.

Before we start I want you to register if you have a firewall, so I can see if they are of any help.

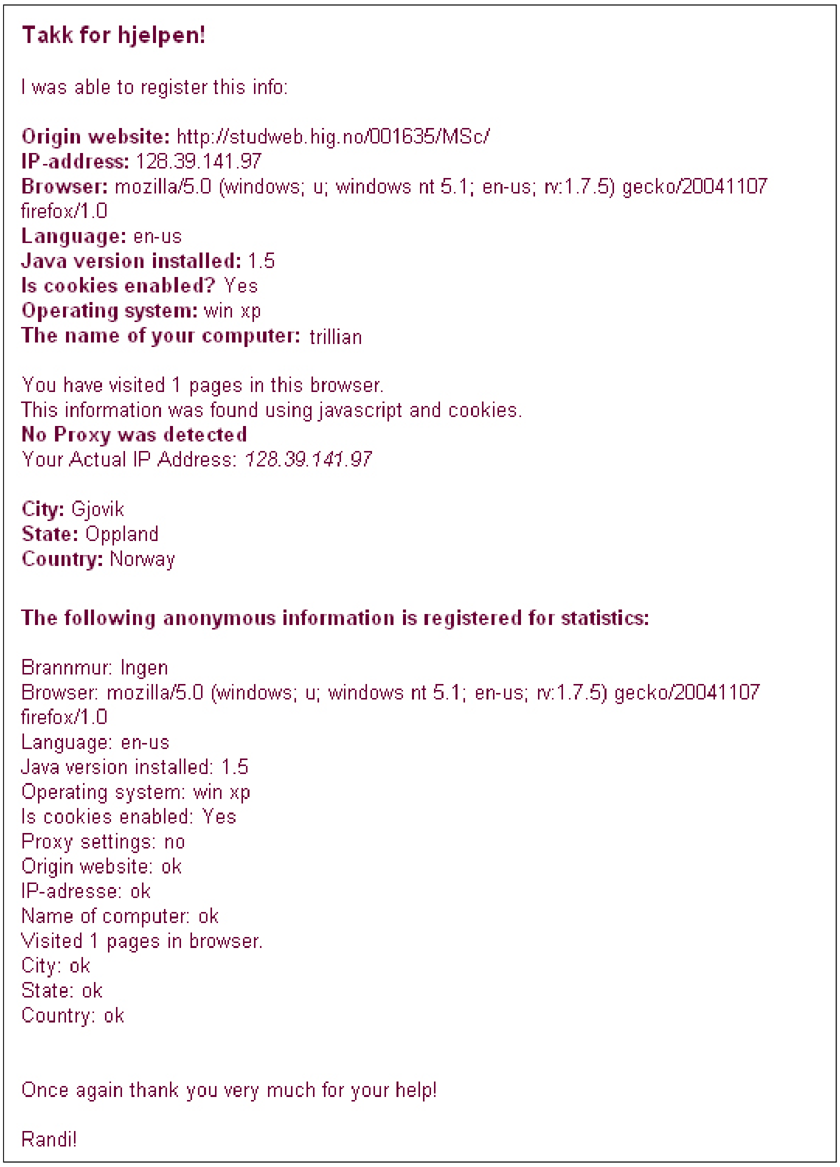
Ingen/None	●
Norton Internet Security / Personal Firewall	○
Norman Personal Firewall	○
Zone Alarm	○
Zone Alarm Pro	○
Panda Internet Security	○
eTrust EZ Firewall	○
McAfee Firewall	○
PC-cillin Internet Security	○
Sygate Personal Firewall	○
Annen/Other	○
Vet ikke / Do not know	○

Mer informasjon || Høgskolen i Gjøvik

Done

Figure 14: *First page as viewed in Mozilla Firefox*

When the participant had read the information and pressed the button, the next page was displayed, in figure 15 as viewed in Mozilla Firefox.



Takk for hjelpen!

I was able to register this info:

Origin website: http://studweb.hig.no/001635/MSc/
IP-address: 128.39.141.97
Browser: mozilla/5.0 (windows; u; windows nt 5.1; en-us; rv:1.7.5) gecko/20041107 firefox/1.0
Language: en-us
Java version installed: 1.5
Is cookies enabled? Yes
Operating system: win xp
The name of your computer: trillian

You have visited 1 pages in this browser.
This information was found using javascript and cookies.
No Proxy was detected
Your Actual IP Address: *128.39.141.97*

City: Gjovik
State: Oppland
Country: Norway

The following anonymous information is registered for statistics:

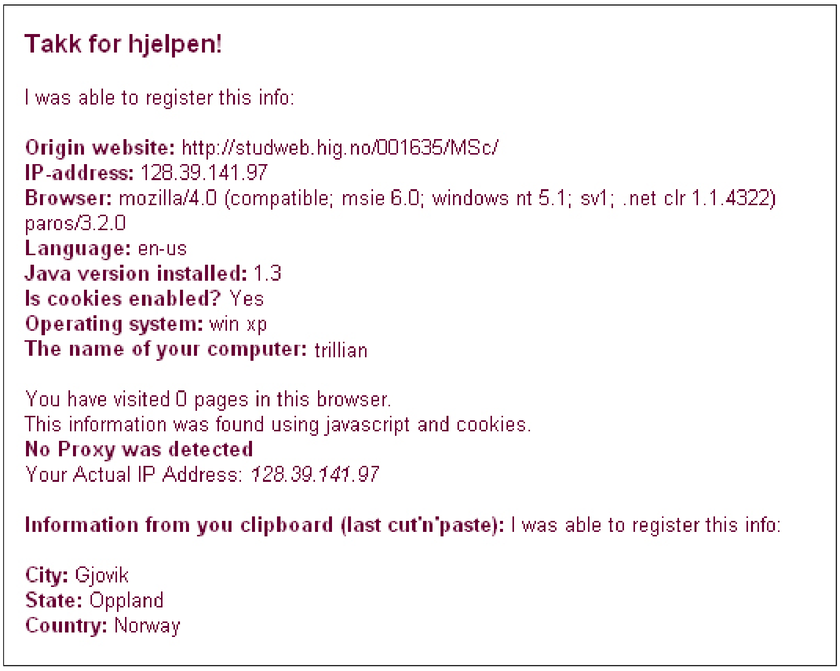
Brannmur: Ingen
Browser: mozilla/5.0 (windows; u; windows nt 5.1; en-us; rv:1.7.5) gecko/20041107 firefox/1.0
Language: en-us
Java version installed: 1.5
Operating system: win xp
Is cookies enabled: Yes
Proxy settings: no
Origin website: ok
IP-adresse: ok
Name of computer: ok
Visited 1 pages in browser.
City: ok
State: ok
Country: ok

Once again thank you very much for your help!

Randi!

Figure 15: *Second page as viewed in Mozilla Firefox*

Figure 16 displays the difference between Microsoft Internet Explorer and other browsers, as the clipboard content is now displayed to the participant.



Takk for hjelpen!

I was able to register this info:

Origin website: http://studweb.hig.no/001635/MSc/
IP-address: 128.39.141.97
Browser: mozilla/4.0 (compatible; msie 6.0; windows nt 5.1; sv1; .net clr 1.1.4322) paros/3.2.0
Language: en-us
Java version installed: 1.3
Is cookies enabled? Yes
Operating system: win xp
The name of your computer: trillian

You have visited 0 pages in this browser.
This information was found using javascript and cookies.
No Proxy was detected
Your Actual IP Address: *128.39.141.97*

Information from you clipboard (last cut'n'paste): I was able to register this info:

City: Gjovik
State: Oppland
Country: Norway

Figure 16: *Second page addition in Internet Explorer.*

Test results:

The results from the laboratory test is shown in the tables following. A *y* means that the information was found and could have been stored, an *n* means that the information was not found. A *dash* (-) means that, in the cookie variable, only a 0 (zero) was found, and it is not possible to know whether this is because a new browser window has been initiated for the test or if the variable has been emptied.

Firewall used	Browser	Java	OS	Cookies	Proxy	Origin	IP	Name	Pages	Geo	Clipboard
Do not know	Firefox 1.0.1	1.5	xp	y	n	y	y	y	y	y	n
Do not know	MSIE 6.0	1.3	2k	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	2k	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	2k	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	2k	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	2k	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	2k	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	2k	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Do not know	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
None	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
None	Firefox 1.0	1.5	xp	y	y	n	y	y	y	y	n
None	MSIE 6.0	1.3	98	y	n	y	y	y	-	y	y
None	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
None	Opera 7.54	1.3	xp	y	n	n	y	y	n	y	n
Other	Firefox	1.5	xp	y	n	y	y	y	y	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	y	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n
Other	Firefox 1.0	1.5	linux	y	n	y	y	y	n	y	n

Table 1: *Test results part 1*

Brannmur	Browser	Java	OS	Cookies	Proxy	Origin	IP	Name	Pages	Geo	Clipboard
Other	Firefox 1.0	1.5	Unknown	y	n	y	y	y	y	y	n
Other	Firefox 1.0	1.5	Unknown	y	n	y	y	y	y	y	n
Other	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
Other	Firefox 1.0	1.5	xp	y	y	y	y	y	y	y	n
Other	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
Other	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
Other	Konqueror 3.3	1.5	Unknown	y	n	y	y	y	y	y	n
Other	MSIE 6.0	1.3	2k	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	2k	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	2k	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	xp	y	y	y	y	y	-	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Other	MSIE 6.0	1.3	xp	y	y	y	y	y	y	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Other	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Norman Personal Firewall	MSIE 6.0	1.3	xp	y	n	n	y	y	-	y	y
Norman Personal Firewall	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
Norman Personal Firewall	Opera 7.54	1.3	xp	y	n	y	y	y	n	y	n

Table 2: *Test results part 2*

Brannmur	Browser	Java	OS	Cookies	Proxy	Origin	IP	Name	Pages	Geo	Clipboard
Norton Internet Security	Firefox 1.0	1.5	2k	y	n	n	y	y	y	y	n
Norton Internet Security	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
Norton Internet Security	Firefox 1.0	1.5	xp	y	n	n	y	y	y	y	n
Norton Internet Security	Firefox 1.0	1.5	xp	y	n	n	y	y	y	y	n
Norton Internet Security	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
Norton Internet Security	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Norton Internet Security	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Norton Internet Security	MSIE 6.0	1.3	xp	y	n	n	y	y	-	y	y
Norton Internet Security	Opera 7.54	1.3	xp	y	n	y	y	y	-	y	n
Zone Alarm	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
Zone Alarm	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Zone Alarm	MSIE 6.0	1.3	xp	y	n	y	y	y	y	y	y
Zone Alarm	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y
Zone Alarm	Opera 7.54	1.3	xp	y	n	y	y	y	y	y	n
Zone Alarm	Opera 7.54	1.3	xp	y	n	y	y	y	y	y	n
Zone Alarm Pro	Firefox 1.0	1.5	xp	y	n	y	y	y	y	y	n
Zone Alarm Pro	MSIE 6.0	1.3	xp	y	n	y	y	y	-	y	y

Table 3: *Test results part 3*