Are the Norwegian Internet users ready for the new threats to their information?

A survey on awareness and use of preventive technologies

Freddy L. Andreassen

Master's Thesis
Master of Science in Information Security
30 ECTS
Department of Computer Science and Media Technology
Gjøvik University College, 2007

Avdeling 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

In a setting where technology, and thus threats to this technology as well, evolves rapidly, it is important for designers of new services utilizing this technology to address to a greater extent the requirements the users of the services put on the system. Most are aware of the threats to conventional valuables, e.g. we do not leave our car unlocked in the street at night and most invest in insurance to protect the value it represents. But are people aware that their electronic valuables, in the form of information, are also prized targets for online criminals?

With services like Mypage(Minside) that aggregate much information and makes it available to the users, we see a shift in the responsibility for protecting this information. We argue that this development could make the uses more attractive to attackers, we are already seeing this with trojans attacking individual users of Internet banking. The stealing of financial information is a direct approach, with the instant transfer of available funds in accounts. But if identity theft and social engineering becomes more common in Norway, we could see users of information aggregating services like Mypage becoming prime targets for attackers looking for information on targets.

If the users are responsible for protecting the information on their computers, it is necessary to put more focus on the users when identifying the system requirements. If the trend of attacking individual users rather than the central servers is continuing, the designers of such systems should make sure that the users are capable and willing to take the responsibility for protecting the information on their own computers. If users are not capable or willing to make effort of protecting their computers, they might choose not to use the service instead.

This thesis investigated levels of awareness towards privacy and threats to an online computer, and in addition the use of preventive technologies such as anti-virus, anti-spyware, firewall and popup-blocker. From the study of previous work on privacy awareness, we found many stating privacy is important, but not putting words into action when it comes to exercising privacy rights or actively protecting their privacy. In addition, recent surveys on malicious software show a very large part of consumer PCs infected.

From our survey and subsequent analysis we confirmed the importance of awareness in explaining use of preventive technologies. We also found the extent of preventive technology use, with both good and bad news. We saw an almost universal use of anti-virus among our survey respondents, but only approximately half of the people asked use anti-spyware.

Sammendrag

I en setting hvor teknologi, og dermed trusler mot denne teknologien, utvikler seg veldig raskt, er det viktig for utviklere av nye tjenester som benytter seg av denne teknologien og i større grad fokusere på de kravene brukerne setter til systemet. De fleste er oppmerksomme på trusler mot våre konvensjonelle eiendeler, man setter for eksempel ikke fra seg bilen ulåst om kvelden og de fleste tegner forsikringer for å beskytte verdien den representerer. Men er folk oppmerksomme på at elektroniske verdisaker, i form av informasjon, er like verdifulle for kriminelle på Internett?

Tjenester som Minside fra Norge.no, aggregerer mye informasjon på ett sted og gir brukerne tilgang til denne. Vi mener dette kan gjøre brukerne mer attraktive for angripere. Vi ser allerede denne type tilnærming hvor trojanere angriper enkeltbrukere av nettbank. Tyveri av kontodetaljer og påloggingsinformasjon er en direkte tilnærming, med umiddelbar gevinst i form av overførsel av tilgjengelige midler. Men hvis identitetstyverier og sosial entrepenørkunst(social engineering) blir mer vanlig i Norge, så kan brukere av informasjonsaggregerende tjenester som Minside bli populære mål.

Hvis brukerne får noe av ansvaret med å beskytte sin egen informasjon, bør ekstra fokus rettes mot nettopp brukerne av tjenesten under identifiseringen av krav til tjenesten. Hvis trenden med å angripe enkeltbrukere fremfor sentrale servere forsetter, bør designere av slike tjenester forsikre seg om at brukerne er villige og i stand til å beskytte informasjonen på egne datamaskiner. Hvis brukerne ikke er i stand til eller ikke er villige til å beskytte datamaskinene sine tilstrekkelig, kan det hende de lar være å bruke tjenesten i isteden.

Denne masteroppgaven undersøkte hvor oppmerksomme norske Internett-brukere er på områdene personvern og trusler mot datamaskiner koblet til Internett. I tillegg undersøkte den bruk av sikkerhetstiltak som anti-virus, anti-spionprogramvare(spyware), brannmur og sprett-opp-vindu(popup)-blokkerer. Fra studien av tidligere arbeid om fokus på personvern fant vi at mange sier personvern er viktig, men lar være å følge opp når det gjelder å benytte seg av personvernrettigheter og å aktivt beskytte sitt eget personvern. I tillegg finner undersøkelser en veldig stor andel av private datamaskiner infiserte med ondsinnet kode.

Fra vår undersøkelse og den etterfølgende analysen fikk vi bekreftet hvor viktig oppmerksomhet på problemene er for bruk av sikkerhetstiltak. Vi fant også store forskjeller i bruk av de forskjellige sikkerhetstiltakene vi spurte om. Nesten alle sier de bruker anti-virus, mens bare omtrent halvparten av de spurte sier de bruker anti-spionprogramvare.

Preface

This MSc Thesis will complete my 2 year MSc degree in information security at Gjøvik University College. My previous degree in economics and IT-management was the reason for choosing a less technical topic for my thesis. I actually started working on a different thesis last fall, but when GUC announced an available thesis in affiliation with a project at the Norwegian Computing Center(NCC), I was immediately interested. I started the work on this thesis in early February.

So this thesis was written in affiliation with the PETweb-project at NCC, more information about the project can be found at http://petweb.nr.no.

I would like to thank my supervisor Einar Snekkenes and the other people that has helped me through the work with this thesis, such as Jan Erik Østvang at KInS, Tore Orderløkken at NorSIS, Åsmund Skomedal and the others from the PETweb-project, my classmates at GUC, and the respondents who took the time to participate in my survey.

Freddy Lønne Andreassen, 26th June, 2007.

Contents

List of Figures

List of Tables

1 Introduction

1.1 Topic

The Mypage(Minside) website at Norway.no(Norge.no) was launched as an online portal for the Norwegian citizens' communication with governmental and municipal service providers. It is to provide a single contact point for online public services, to simplify communication with public bodies and to provide the individual with information on what is stored about him/her in public registries.

The topic for this thesis is knowledge and attitudes relating to security and privacy, among Norwegian Internet users; the potential users of Mypage.

1.2 Keywords

Privacy protection, privacy awareness, threat awareness, use of preventive technologies.

1.3 Nøkkelord

Personvern, fokus på personvern, fokus på trusler og angrep, bruk av sikkerhetstiltak.

1.4 Problem description

An information portal like Mypage aggregate a lot of personal data and when making it available to the individual user through the portal, the information is placed in the user-computer (if only temporarily). We suggest that this will give attackers greater incentive for attacking individual users rather than the centralized locations, such as the service providers or the portal servers.

We are already seeing this development in for example the case of trojan horses stealing banking information and performing money transfers, when users use Internet banking. This is the attack approach that gives an instant financial benefit and therefore is likely to be a primary choice for attackers. But identity theft, which could be the result of obtaining lots of personal data on a person, would become more attractive should the online banking approach become more difficult to pursue successfully.

Such a future development would be a potential escalated threat to privacy and it is therefore important that the users of Mypage.no are protecting themselves against these types of attacks.

In the user requirements survey done by Vindfang[18] for Software Innovation before the Mypage portal was implemented, several suggested requirements was made based on the feedback from potential users. And although the survey did not include questions on security, 36% of the 509 respondents mentioned security as a worrying aspect of implementing the portal.

This gives one reason to think that users are aware of the security issues about concentrating personal data and making it available for the individual at the individuals'

location. But do users know about threats to their privacy online? And do they have the competence and will to do what it takes to protect themselves?

1.5 Motivation and justification

To know how well the users are protected and what they are willing to do to protect themselves is important to all that design web-based information services. When placing some of the responsibility for protection on the users, it is essential to investigate if the users are capable and willing to accept this responsibility. If users are not capable or willing to make effort of protecting their computers, they might choose not to use the service instead.

1.6 Research questions

From the work with the state of the art we formulated the following research questions:

1. What is the awareness on the issues of privacy and threats among the Norwegian Internet users?
2. To what extent is core preventive technologies utilized by Norwegian Internet users?
3. How does awareness affect the use of preventive technologies?
   - Does any specific knowledge affect the use more than others?
4. Is there a level of awareness that triggers the use of preventive technologies?
   - Are there different levels of awareness ”needed“ for adoption of the different preventive technologies?
5. How interested are potential users of web-based services like Mypage in educational material and security measures, if made available in affiliation with the Mypage portal?

2 State of the art and related work

This chapter will present the issues at hand, beginning with privacy and privacy awareness. We then move on to threats to privacy and threat awareness before finishing with preventive technologies adoption. The purpose of this chapter is to give us the theoretical foundation for our survey, to make us able to gather the right data for answering our research questions. The main areas of theory was chosen partly based on the results and ideas of Freeman and Urbaczewski[14], where they argue that privacy, more than performance, is what makes people act against spyware.

Glossary and translations

Towards the end of the report, glossary and translations are available. At first occurrence of a translated text, the original Norwegian name or word(s) are included in a parenthesis.

2.1 Privacy

In this section we will give a short introduction to relevant privacy issues. We will start off with defining privacy and then link the privacy issue to the Mypage portal.

2.1.1 What is privacy?

Unfortunately, the concept of privacy is not something that can be narrowed down to a single value. There are several intertwined ideals, views and interests that need consideration when explaining privacy. These concepts are also weighted differently and has different meaning in different countries and societies and has in addition changed quite a bit from its origin.

The beginning

The concept of privacy or at least the systematic discussion of privacy is said to be introduced in 1890, by Samuel Warren and Louis Brandeis in the article “The right to Privacy” in the Harvard Law Review[56]. They argued for “the right to be let alone” and how the law at the time supported that right.

Later, in 1967 it was described by Alan F. Westin[58] to be the right to determine how information stored about us are spread. This includes when, how and to what extent that information about us is forwarded to others. It is said that this book of 1967, started the modern international discussion on the concept of privacy. Westin defined the term privacy as “the claim of individuals, groups or institutions to determine for themselves when, how and to what extent information about them is communicated to others”.

He continues with describing four states of privacy; solitude, intimacy, anonymity and reserve:

Solitude is the separation from the group, where the individual can be freed from the observation of other persons. This is the most private an individual can get.

Intimacy is when the individual is part of a small group, where it is accepted that the

group practice seclusion. This could be husband and wife, family, friends or close work clique.

Anonymity occurs when the individual seeks and finds the freedom from being singled out and identified in open spaces or in a crowd. This is necessary for relaxation and freedom in such places, that one can expect not to be personally identified.

Reserve is about the individuals selectiveness on what information it wants to communicate to others, and that others respect this reservation from communicating certain information. This is the most subtle form of privacy.

These states of privacy performs several functions for the individual in a democratic society. These functions of the individual privacy can be described as personal autonomy, emotional release, self-evaluation, and limited and protected communication:

Personal autonomy: Each individual consists of a "core self", that should not be affected or manipulated by others. This to establish a sense of individuality for the individual and enable the individual to develop this individuality and to enjoy the individual choice.

Emotional release: When interacting with its surroundings, the individual plays many roles depending on with what and who it is interaction and who is observing this interaction. But the individual cannot play these roles continuously for long periods of time, this would make the individual stressed and would ultimately destroy the human organism. Therefore he/she needs emotional release from the roles, to just be one self and do what it pleases.

Self-Evaluation: The individual needs to make decisions in private, it needs time to process and reflect on information unaffected by others and take action according to its own preferences.

Limited and protected communication: If all were to speak and do exactly what we wanted at all times, much structure of social life would collapse. We all filter what we communicate to others and this is necessary to enable interaction between different individuals.

Early Norwegian work and laws concerning privacy

A note on legislation references in the following chapter; most of the English translations of the legislation have been collected by The Faculty of Law Library at the University of Oslo and are available for the public at their website[54]. The collected translations have been done on the initiative of Governmental Agencies, Royal Ministries and private institutions, but they are unofficial and only updated at the time of the translation.

We will keep the focus on the Norwegian discussion on defining privacy. Already from 1902, the Norwegian General civil penal code[32] (Den almindelige norske straffelov[24]) included the Â§390 who states punishment for "violating the private life by publicly announcing private or domestic matters" (translated from Norwegian). This law was actually based on an addition, from 1889, to the previous law on punishment for criminal behavior of 1842.

But the earliest work on describing the concept of privacy in Norway started with the

book of Blekeli and Selmer[5], who in the 1970's introduced the theory of interest. In the book, Blekeli starts off by stating that privacy could be thought of as the interest of the individual to exercise control on the information describing him/her.

He continues with splitting this interest into three sub-interests, these are described as the interest of discretion, the interest in completeness of information and the interest of involvement:

Discretion. This interest is two-part; First is the interest is about control of the propagation of information about the individual. An example is information about an individuals criminal record, when the debt to society has been paid and the sentence for the criminal activity has been served. The second is processing and use of information once it has been spread; the individuals control of what use the information is put to by an organisation after it is recorded.

Completeness. The individuals interest in the completeness of information, upon which decisions regarding the individual are being made. An example is that all relevant information is considered when governmental decisions are being made about the individual.

Insight. This is the individuals interest in involvement in decisions being made regarding him/her and what information this information is made upon. This is a prerequisite for the previous interest.

In the same book, Selmer continues with the discussion of these individual interests versus the interests of the society. The social interests of national security, public safety, cheap and effective public administration, and so forth are interests that must be weighted against those of the individual.

Selmer describes what he sees as the four privacy interests of an individual and use these as the starting point for discussing social interests that must be weighted against an individuals interests:

Discretion. A functioning society needs information on individuals for e.g. taxation, social welfare, population statistics etc. A conflicting area of individual and social interests is the rationalisation aspect of reusing information. Large savings in resources can be made from combining and distributing informations to several governmental bodies or others, making decisions on the same individual.

Involvement. The same argument is used against the individuals interest in being involved in decisions made abut him/her. If we enable too much involvement of the individual in each decision, it would make the administration very costly and little effective.

Citizenfriendly administration. The need to rationalise can also conflict with an individuals interest in citizenfriendly administration. Glitches in automated processing of information can lead to all sorts problems for an individual.

Avoiding excess control or untitled use of authority. The individuals interest in avoiding excess control or untitled use of force is not directly threatend by centralisation

of information and rationalisation, but the possibility of using the information for such purposes is much greater once the centralisation has been done.

Then in 1978, the Act relating to Personal Data Filing Systems, etc.(Personregisterloven), was sanctioned and it came in to force from 1.1.1980 (in Norwegian[25] and English[30]). This law stated that a data inspectorate should be established, it defined certain requirements for all personal data registers, it required computerized registers and registers with sensitive information to have concession from the Data Inspectorate(Datatilsynet) and it regulated certain types of businesses such as credit reference and direct marketing companies.

Current privacy discussion

The theory of interests is followed up by Jon Bing[4] in 1991, when writing the booklet "Privacy in the danger zone"(translated from Norwegian). He discusses several principles of privacy as well as give his own view on the interests model of Blekeli and Selmer[5].

He discusses the lever-principle, that the interest of controlling information about one self sometimes must give to other interests. These could e.g. be the protection of life, health or common interests. This has been a much debated subject since the 9/11 terror attacks.

Bing continues with describing the power relations in the society, e.g. those between government and citizen, between employer and employee and the one between the individual and its local community. To keep the balance of power "equal" and not tip the weight-scale to much in one direction in these examples, the interests of all parties needs protection.

He then describes the 3 interests of the individual as the interest of discretion, completeness and visibility. The interests are basically the same as those of Blekeli and Selmer[5] and are explained as this; discretion is the interest of controlling the collection of information, the purpose and use of the information and the further distribution of the information, completeness is that decisions made on basis of personal data, should consider all information and not only parts of the information. Implies also requirements to quality of the information and visibility is that the individual has a right to be informed of what information is stored on him/her, how it is processed, the purpose and how it is distributed, upon request.

Further he describes the 3 public interests as the interest of a citizen-friendly administration, a robust society and the level of society surveillance. The interests are explained as this; a citizen-friendly administration means that more readily available and structured personal data would make the contact with government more individually adapted in addition the quicker processing of requests and applications, a robust society is that better control from the government gives it a greater chance of avoiding that public services becomes unavailable and the level of society surveillance implies that widespread analysis and connection of registers would enable better control of unwanted elements in a society, such as criminal behavior.

Current laws

In 1997, a committee delivered a report to the government entitled "A better protection of privacy"[37]. This report was part of the preparatory work done before introducing the Norwegian Personal Data Act(Personvernloven) we have today(in Norwegian[26] and English[31]).

They defined privacy from 3 types of perspectives and 7 interests. The perspectives were on integrity, power and decisionmaking. The integrity perspective is about the private sphere around an individual, much the same as the early definitions of privacy. The power perspective is the same as described by Bing, about power relations between groups or individuals in society. Last, the perspective of decisionmaking is about how decisionmakers in society, makes these decisions. The decisions range from whether or not a bank gives us a loan or that we are granted welfare support from the government. Common is that these decisions affects us as individuals and this perspective is to prevent mass-decisionmaking done by automated systems.

They describe the same 6 interests as Bing, but includes a seventh one for the individual. This is the interest of respect for the personal life of an individual. This interest referring to the "right to be let alone" and arose to the issue of direct marketing.

Making the concept more practical oriented

Dag Wiese Schartum and Lee A. Bygrave[46] acknowledges the same three perspectives as described by the committee above. But they argue that the theory of interests is a bit vague, when one is about to put the theory into practice. They argue for a set of requirements to be used instead, that represent the intention of the traditional interests.

This attempt to make the theory more practical resulted in 18 requirements on the 5 areas of deciding accessibility of data on one self, access to and knowledge of data, data and processing quality, reasonable control and userfriendly processing.

1. Interest on deciding accessibility of data on one self:
    - Established relationship of trust.
    - Confidentiality.
    - Privacy protection.
    - Personal integrity.
2. Interest on access and knowledge:
    - Information on laws and rights.
    - Information on general access.
    - Information on individual access.
    - Justification of access.
3. Interest on data- and processing quality:
    - Quality of data.
    - Quality of processing.
4. Interest on reasonable control of processors:

- Accordance between guidance and control.
- Accordance between pre-control and post-control.
- Accordance between control in favour and in disfavour of the registered.
- Accordance between external and internal control.

5. Interest of userfriendly processing:
   - Ease of making the users opinions heard.
   - Ease of understanding.
   - Ease of dialog.
   - Stability of systems and routines.

2.1.2 Privacy and new technology

> "New technology has in it self neither a positive nor negative impact on privacy. It is how we make use of that technology and how we set the terms for development of new technology, that decides how the technology will affect privacy."

A quote from the annual report of Data Inspectorate in 1995, found in the government report of 1997[37]

Schartum and Bygrave[46] points out that the technology enables a much greater extent of processing personal data. New technology gives the possibility of new forms of personal data, such as images, video and sound. In addition, the technology enables many more ways of recording this data and also the increased automation of this recording.

This has been followed up by laws and regulations on privacy. Before 1970 the focus was on not intruding into the private sphere of the individual and on not violating the honour of an individual. After 1970, the focus has been more directed towards the computerized processing of personal data and personal data registers.

Beth Givens[16], director of Privacy Rights Clearinghouse argues that electronic records increases the effects of mistakes. That small errors can destroy reputations and ruin lives. It may also cause people to withdraw from the public life if it means having information on them published. Her paper is based on the practices of government bodies in the US.

But we have several examples of personal data being published in Norwegian media lately, where even sensitive personal data has been mistakenly published online. This includes Mandal, Ålesund and Elverum municipal governments, the county administrator of Vest-Agder, the University of Oslo, BNbank ASA and others. See references to articles in Appendix A.

2.1.3 Summary

The term privacy has changed quite a bit from its origin. From the private sphere model of the early nineteenth century to the lever model with perspectives and interests we have today. We have moved from privacy as an ideal, via interest models and now the work on developing a framework for better understanding privacy. It seems more and more nuances of the term privacy have arisen as the work has progressed, and thus making it more and more complex.

The progress of technology has also forced changes to the laws and regulations. Informa-

tion technology has made the processing of personal data easier and more extensive, and therefore more vulnerable to mistakes and the like[16]. Thus, it appears that it is not the technology itself that threatens privacy, it is our use or rather misuse of technology that is the real threat to privacy[37].

In the following sections we will sum up work done on privacy awareness and then present the most common threats to users privacy on their own computers.

2.2 Privacy awareness

Several surveys have been conducted in the recent years, to investigate the populations knowledge and attitudes regarding privacy and what they know of threats, rights and legislation. We have looked mainly at results from the population studies of the Institute of Transport Economics(Transportøkonomisk Institutt)[43], the Norwegian Board of Technology(Teknologirådet)[35], the Norwegian Computing Center(Norges Regnesentral) [51] and the European Commission[10]. The latter not including Norwegian citizens, but includes results from Sweden, Denmark and Finland. We will only include these Scandinavian results from the European Commission survey, and compare them with the Norwegian studies.

In addition, we have looked at the subsequent company studies from the Institute of Transport Economics[42] and European Commission[9], that were conducted subsequent to each of the population studies. Also, some conclusions from the user requirements survey[18] and Urban Eye[53] affiliated surveillance surveys from Oslo[45] and Berlin[20] where considered to get the best possible picture of what the Norwegian population thinks and knows about privacy.

These abbreviations for the surveys are used in the following discussions:

Privacy - Citizens:

ITE1	Institute of Transport Economics citizens[43]
NBT	Norwegian Board of Technology[35]
NCC	Norwegian Computing Center[51]
EC1	European Commission citizens[10]

Privacy - Companies and organisations:

ITE2	Institute of Transport Economics companies[42]
EC2	European Commission companies[9]

Other - Citizens:

SI	Software Innovation[18]
UE1	Urban Eye survey Oslo[45]
UE2	Urban Eye survey Berlin[20]

2.2.1 Technical information on the surveys

ITE1

Was conducted in 2005 and is a representative survey for the Norwegian population aged 15 and above. 1000 individuals were interviewed by telephone. The main focus was on

attitudes towards privacy, knowledge on what situations and actions cause personal data to be collected and processed, and knowledge on rights and legislation.

NBT

The NBT survey was conducted in 2004, and included interviews of 48 Internet and mobile phone users in 6 focus-groups. 4 groups represented youth between 17 and 19 years old and 2 groups included adults aged between 30 and 40. The main focus was on the users views on electronic tracks and privacy.

NCC

This survey had two parts, one conducted during the winter 1999/2000 and one during the winter 2000/2001. The postal survey had 5660 respondents in 99/00 and 5376 respondents in 00/01, representative for the Norwegian population. The focus was on what information the respondents would give away in an on-line shopping situation.

EC1

The survey EC1 was conducted in 2003, in the 15 membership countries of the European Union. In total, 16124 respondents aged 15 and above was interviewed face-to-face in the appropriate national language. Main focus of the survey was on citizens views about privacy related to information stored about them in companies and organisations.

ITE2

Conducted subsequent to the population study, this study was also done in 2005. 424 companies and organisations with 4 or more employees replied via an Internet questionnaire. Focus was on what personal data was processed, attitudes and knowledge towards legislation and authorities and weighting of privacy interests against other interests.

EC2

This survey was also conducted subsequent to the EU citizen survey, in 2003. A total of 3013 companies and organisations with 20 or more employees from the 15 membership countries was interviewed by telephone. Main focus was to investigate data protection awareness and knowledge of the legislation by the data controllers.

SI

Conducted by Vindfang AS on behalf of Software Innovation in 2005, before the Mypage.no project. 509 respondents representative for the Norwegian population of Internet users aged 18 and above. Focus was to investigate the users requirements and preferences related to the Mypage.no portal.

UE1

This survey was conducted in 2004 and included a quantitative part and a qualitative part. 218 was recruited in downtown Oslo for the quantitative interviews and 13 in-depth interviews were conducted. Focus was on how invasive and protective urban dwellers thought Closed Circuit Tele Vision(CCTV) was.

UE2

Conducted in 2004, this survey consisted of a quantitative and a qualitative part. 203 people was interviewed outside shopping malls and then 10 of these were interviewed in-depth afterwards. Main focus was on what people in urban areas thought and felt about CCTV.

2.2.2 Importance of privacy protection

The EC1 survey found that 59% of Scandinavians were fairly or very concerned about the protection of privacy and in the ITE1 survey, 90% of respondents stated that we need a strong data inspectorate. Also, the NBT survey report states that most people care about the privacy of the individual, but many, and especially the youth, were not able to pinpoint why. Most people stated that they did not ponder during everyday life, but considered worrying as they were presented theoretical scenarios of misuse.

The NCC survey found that the Norwegian Internet users consisted of 9% users very concerned about their privacy and 29% marginally concerned concerned of their privacy. The majority, the remaining 62%, were somewhat concerned. Respondents were grouped according to how much personal data the respondents would give away to shop online. Notable is that these results are from the selection of people already shopping online.

The SI survey indicated a possibility of Mypage.no users being rather concerned with the security of their personal data, when accessing them via the portal. Although no direct questions where asked on the subject, 36% of the respondents said that security was an important concern, when asked of any negative issues of such a portal.

These results indicate that citizens are at least fairly interested in privacy and think these are important matters. But one contradictory result is that the EC1 survey found 65% of Scandinavians agreeing to privacy awareness in their home country being low.

2.2.3 Knowledge

The ITE1 survey found the following numbers from the Norwegian respondents; As many as 68% on average, said they knew of the right to access information on them selves, the duty of the data controller to inform the registered about what information has been collected and for what purpose it was collected, the right to deny the use of personal data for marketing purposes and the need for the data controller to get consent from the registered in some cases.

But one interesting aspect is that e.g. 84% of Norwegians has never exercised the right of access to information about themselves, to see what information is stored about them at a data controller. Of these, 68% say it is either because they have not reflected on the fact that information is being gathered or that they simply do not care. It should be noted that the question was asked the whole selection, not just the ones that knew of the right. Thus are the 32% that did not know of the rights probably well represented in the 84% that never exercised their rights.

The EC1 survey found lower numbers for our Scandinavian neighbours; only 46% on average had heard of the four rights described above. Only 11% of the 26% that knew of the right of access, had ever exercised it. An independent authority, such as the Data Inspectorate in Norway, exists in all Scandinavian countries. This authority monitors compliance with current laws and regulations, and this was only known to 30% of the respondents in the EU1 survey. The same result for Norwegians in the ITE1 survey was 44%. But of these 44%, only 33% could name the authority as the Data Inspectorate.

Above many states privacy to be important and this is supported by the rather large number of people knowing about the various rights and duties in the legislation. But when we

look at the low percentage of people that know the name of independent authorities or the reasons people give for not exercising their rights, it seems not many follow through on their initial statements about importance of privacy.

2.2.4 Trust in data processors

Both the ITE1 and the EC1 surveys concluded that citizens trust many data controllers highly. From the other Scandinavian countries the health services and doctors (89%), police (84%) and banks (82%) were most trusted. Organisations or companies most often stated as untrustworthy was mailorder companies (65%), non-profit organisations (44%) and credit reference companies (38%).

Similar results were found in Norway, where health care services and police (91%), banks (87%) and many other public services enjoying trust from more than 80% of the citizens. Companies and organisations like telecom companies (55%), companies administering tollbooths (51%) and non-profit organisations (46%) are most often stated as untrustworthy.

The reasons to these numbers are puzzling, it seems to be mostly “positive” companies and organisations, meaning the companies and public bodies there to our benefit, that are enjoying the most trust. Similar, the companies we trust the least, are in general companies and organisations we perceive as “negative”, in that they e.g. sends us bills and collects money from us. Comparing public bodies and private companies in the ITE2 survey, show not much difference in compliance with legislation at the two types of organisations. This gives us reason to think that the trust of the citizens might be misplaced.

2.2.5 Trust in legislation and compliance to legislation

The ITE1 survey found people having great trust in privacy protection practises in Norway. They trust in the compliance to legislation to the extent that 54% say it is perfectly safe to give away personal data. 78% also agree to that the Data Inspectorate makes sure that no one misuse collected information. 76% do not think the current legislation is to harsh.

In the EC1 survey too, a high percentage (67%) said that national laws protect their privacy to a high degree. But the EC1 survey also found that 50% of Scandinavians do not think that the current legislation is able to cope with future developments of privacy issues, in regard to Internet use. Notable is that an additional 25% do not know whether or not it is adequate for the future.

Here we see that people have great trust in legislation too, but some are worried about future developments in the privacy issue. But it seems the worry is connected to the use of Internet, which we also see in the ITE survey, where 67% say they are careful with what information they give away on-line. This is also supported in the NBT survey, where most trusted the government and large private companies, but where rather sceptical towards Internet-based companies. The NCC survey did, however, show that only 9% was among the very concerned, but this could perhaps be somewhat explained by the fact that only online shoppers were included in that section of the survey. The very concerned are probably not well represented among the online shoppers in the first place.

2.2.6 Views on surveillance

From the UE1 and UE2 surveys we conclude that people are rather positive towards CCTV surveillance. In Oslo, places like in banks (91%), in shops (85%), railway and subway platforms (84%) and in taxis (72%) are most accepted for surveillance. In Berlin, bank counters (86%), railway and subway platforms (85%), shops (69%) and in open areas of shopping malls (61%) were the most accepted places to have surveillance cameras.

There is a more negative attitude towards surveillance of more intimate places, such as fitting rooms in stores and changing rooms in sports centers, but still 21% of Norwegian respondents are positive to cameras in store fitting rooms.

2.2.7 Privacy versus other interests

We have the following numbers from the ITE1 survey; 72% think the Data Inspectorate should pay more attention to other interests than privacy. 80% of Scandinavians totally or partly agree to that the police should keep the Internet under surveillance for suspicious behaviour.

From the EC1 survey, 68% agree totally or with some restrictions that individuals should agree to having their telephone calls monitored in the fight against terrorism. 69% say the same about Internet surveillance for the same purpose.

These are also troublesome numbers, as it seems both Norwegians and the rest of Scandinavians are not very vigilant about the privacy interest when put up against more public interests.

2.2.8 Attitudes and knowledge in companies and organisations

The ITE2 survey shows that, although a general positive attitude is found among the companies and organisations, there is little knowledge of legislation and not much compliance with it. 70% say we need the Personal Data Act to ensure privacy protection and 59% say we need a strong Data Inspectorate. But only 16% say they know the Personal Data Act well and the same percentage say they know what role and tasks the Data Inspectorate performs. Only 4% say they follow all requirements from the legislation and when asked why not all follow the requirements, 74% states lack of knowledge about legislation as main reason. Also, not much difference is found between private/public companies and public bodies and service providers in this matter.

From the EC2 survey, we find similar numbers for the Scandinavian companies and organisations. 88% think privacy legislation is necessary and 53% think their country has greater compliance with the legislation, than other countries. But 52% say lack of knowledge of legislation is main reason for not complying with the requirements.

2.2.9 Summary

The results seem at first somewhat contradictory. The Norwegian and Scandinavian citizens say that the general privacy issue is important to them and most are familiar with rights and duties stated by the legislation. But we do not have much knowledge of authorities on the matter and we trust most companies and organisations to comply with the legislation. The citizens trust the legislation to protect them and the Data Inspectorate to ensure compliance with legislation. A note is that those who do NOT think legislation protects their personal data, do not exercise their right to access to data about them any

more than those who trust the legislation.

We cannot avoid the discussion on how naive the citizens are on the issue of privacy. In the ITE1 survey, 86% of Norwegians agree that only individuals with criminal intent has reason to dispute camera surveillance. This number is 67% from the UE1 survey and 70% in the UE2 survey. Also, in the NBT survey, statements like "but I do not write that sensitive e-mails" and "I do not have that kind of secrets [about the need for encrypting e-mails]", suggest that people do not see the big picture of privacy. In addition, only 17% had ever heard of tools for limiting their tracks on-line and only 6% used them. This is also backed up by numbers from the EC1 survey, where only 12% use these tools. Common for users that do not use them, is perceived difficulty of installing and using such tools.

It seems then, that perhaps people have been affected by socially accepted attitudes towards privacy. That many say these things are important and that they say they know of legislation and their rights, but that the majority are not willing to do something to actively protect themselves or be vigilant about their own privacy. This is reflected for instance by the small number of people that exercise their rights and the reasons they give for not doing so.

We concur with the final statement in the ITE1 report:

> "If one wishes to strengthen the position of personal privacy in the society, it is problematic to leave more of this responsibility to the individual."

2.3 Threats to the end user

In this chapter we will look at previous work in the field of end user threats. The focus will be on threats most common and most likely to be a result from bad surfing habits and unwise actions from the users. Also we will look mainly at the automated attacks, as they are the most common.

Attacks aimed at end users are very differentiated in approach and techniques used. They range from eavesdropping attacks that try to sniff information without being detected, via types of phishing attacks that tries to deceive the user in some way and to trojans and the like, that attempt infiltrating the user computer without the user knowing.

2.3.1 Browser eavesdropping

Browser eavesdropping is basically sniffing what information your browser discloses about your computer. If a rogue website is configured to sniff as much as possible, quite a lot of information is available from the browsers.

Randi Gjerde[17] did in her masters thesis in 2005, several experiments into what information browsers could leak if a malicious website wanted to record information about the visiting users. Examples of information she was able to extract from test subjects surfing a test website:

- Referer header.
- IP address.
- Browser name and version.

- Java version.
- Part of browser history.
- Operating system.
- Name of computer.
- Clipboard content.
- Geographical location(from IP-address).

Several of these properties could, alone or by inference, be enough to identify a person. This information could also be used for further attacks, as the information will indicate what types of security holes could be present on the computer. It is also worth noting that she did not find it possible for the users to detect if information was recorded or not.

We see the clipboard content as perhaps the most important direct extractable information on this list. It is then worth mentioning that this information was available from all users surfing with the Microsoft Internet Explorer 6.0, but was not available from users surfing with any other browser in the test. These included Firefox 1.0, Opera 7.54 and Konqueror 3.3.

2.3.2 Phising and pharming

Phishing[38], characterized by the use of spoofed messages to lure from users their electronic identities. Commonly used is spoofed e-mail messages warning about e.g. security threats. The e-mail then contains an attached legitimate-looking security update patch or link to one. But if the receiver clicks or runs the patch, malware is often installed instead. Then there is the evolvement into pharming[39], which manipulates the name lookup process that is used to connect to hosts or services. This often exploits security holes in domain name lookup systems, enabling the use of legitimate web addresses and still routing victims to malicious sites.

2.3.3 Malicious software

There are many definitions of malware and its subtypes. But according to Zaytsev[59], the traditional categories of malware are:

Virus Software that infects other software. Recognized by their ability to insert the body of the viral code into the body of the software it is infecting. The goal is to gain control of the infected program. Removing viruses can range from easy, just by removing certain files, to hard when one needs to re-install because the virus has encrypted parts of the infected code.

Worm This type of malware does not infect software, but copies itself and send itself to all computers affected by the weakness the particular worm exploits. Removal is usually easy, as removal of worm components are usually sufficient.

Trojan This malware does not infect or re-produce like the worm. The Trojan runs hidden processes, potentially disclosing personal data, destroy data or interfere with other applications. To remove a Trojan, it needs to be detected and have its files removed. This could be hard, if the Trojan has encrypted or obfuscated its code.

Adware/Spyware These malware types are by definition not “harmful”, in the meaning of destroying files or causing too much hazzle. But they will slow the computer down and generate network traffic. They may snoop information, track user activities, conceal their existence and actively protect themselves from deletion.

Of the wide variety of functions of malware, it is the keylogging and rootkit abilities that that we would like to mention as the perhaps most important. The keylogger part of a piece of malware can log running processes, generate screen shots, track clipboard contents and log keystrokes. The rootkit function is basically a set of tools designed to maintain control of a computer or system after an attacker has gained control over it. This could for instance be by creating a backdoor to the system, giving the attacker the possibility of logging in as he pleases. These tools often include measures for hiding tracks and rootkit presence from the operating system on the infected machine(s), by encryption or obfuscation.

A note to these definitions is that a specific piece of malware may contain components from all groups. A Trojan horse can be the viral body of a virus, along with worm-like code that replicates itself when certain requirements are met in other computers. The only thing positive about this, is that the more complex the malware, the more detectable parts it have. E.g. a virus program might trigger on the same program as an adware remover and so forth.

Looking at the last year of development in malware, we see that the functionality of malware is becoming more diverse and advanced. Earlier, programs more often had one function and thus they were easier to classify. A summary is found in a recent publication from IBM Internet Security Systems; the X-Force 2006 Trend Statistics[22]. It states that the traditional categories of malware are more or less useless now, as most malware now has code belonging to several, if not all, of the traditional categories Zaytsev[59] listed.

The X-Force team[22] describe a new set of classification for malware in 2007, based on primary function of the software:

Worm - Self-propagating software.

Backdoor - Enabling an attacker unauthorized access to a system, .

Virus - Infecting and damaging host, but do not propagate.

Password stealer - Software designed to steal login credentials.

Downloader - Simple software that downloads more advanced malware, once on the target.

Keylogger - Stores all key strokes, for later retrieval.

Dialer - Makes unauthorized connections, with modem connections, either back to attacker or to high-cost services.

Trojan - Appear legitimate, but is hostile and installs hidden code.

Miscellaneous - All other malicious software.

2.3.4 General methods of infection

The most common methods of getting different malware onto a target computer, is by including them in other software. This could be ActiveX components, cracked programs or other types of freeware such as P2P software, plugins, toolbars or other browser helper objects. Other methods include pop-ups, ads, links, e-mail or attachments.

One important note is that, almost always, some form of active involvement from the user is needed. But this can be as little as just clicking a link or pop-up. Some forms of spyware can install themselves just by the victims surfing on e.g. a webpage, but then it is often a matter of exploiting security holes in software. This is often referred to as drive-by downloading.

2.3.5 Subsequent consequences

Although the consequences of theft of personal data are not as direct as e.g. theft of information when accessing online banking services and such, they can lead to e.g. identity theft. Follow-up attacks, like types of social engineering are a very likely consequence should any of the above attacks succeed in disclosing personal data.

Kevin D. Mitnick describes social engineering; “Social engineering uses influences and persuasion to deceive people by convincing them that the social engineer is someone he is not, or by manipulation. As a result, the social engineer is able to take advantage of people to obtain information with or without the use of technology,”[33].

Also, if the information obtained is of sensitive enough character, even attempts of extortion could be a possible consequence.

2.3.6 Other issues

There is the possibility of lost or stolen hardware. According to Privacy Rights Clearinghouse[41], there has been approximately 155,000,000 records containing sensitive personal information involved in security breaches in the US, since they started listing this type of data breaches in 2005. This vast amount should indicate that this might be a problem also in Europe and Norway.

Cheaper hardware and broadband connections present problems as well opportunities to the common internet user. If not protected, todays average PC is a powerful tool to attackers if they gain control over it. Botnets or zombienets is several PC’s and up to several million, that are controlled by a central server. These nets can be used to launch denial-of-service attacks, to send SPAM or even to host illegal content such as child pornography. Examples from Europe[27] and here in Norway[44].

A wireless router is very convenient for sharing an Internet connection in a house, but an unsecured wireless network could be easily exploited since the range of the router is often much larger than the building. All network traffic can then easily be sniffed and the network and Internet access or computers can be exploited and used for a number of criminal activities.

Security messages from browser can also be a problem for common users. Many believe that as long as the browser shows a padlock or similar icon, they are safe from attackers. But this security symbol is only guaranteeing the connection to the closest server at best.

What it does not guarantee is that the server uses TLS or SSL(which are ways to secure the connection from the browser to the server), when or if it needs to forward the traffic to another server.

Another notable issue is the fact that it is not easy to verify if certificates are good or not. Most browsers today warn the user when accessing a website that has got a certificate that does not match the web address. But it is not easy for a common user to distinguish a fake certificate from a real one. We have several examples of sites that have a certificate that seems wrong, one example is the certificate of www.altinn.no, which states it belongs to Accenture. [UPDATE per 29th June 2007: This has now been fixed and the certficate on www.altinn.no now belong to Brønnøysundsregistrene, which Altinn is a part of.] This example was probably just a mishap somewhere, but the point here is that one can easily fake a certificate and it is not easy for the common user to expose this fraud. Also, the tendency is that with many such warnings, we pay less and less attention to them. After a few of these certificate warnings, you might just click on "yes" to make the message go away. And if you accept this certificate, you in fact choose to trust the server and all messages from it.

2.3.7 Scope

So, how real are these threats? Unfortunately we have not found much independent research on this issue, but several software companies, especially in anti-virus and anti-spyware business, present annual or quarterly reports on vulnerabilities and scope of threats.

Vulnerabilities

IBM Internet Security Systems publishes a quarterly threat overview. The Q4 2006 publication states that X-Force analysts checked a total of 7247 vulnerabilities in 2006[23]. This is a 39,5% increase from 2005 and as many as 1230 of these were categorized as high or critical vulnerabilities. 88,4% of these vulnerabilities could be exploited remotely and 52,5% of the vulnerabilities were classified as enough to give an attacker access to a system, in a worst case scenario.

Spyware infections

The annual Online Safety Study done by America Online(AOL) and National Cyber Security Alliance(NCSA) from 2005[2], showed that 81% of US computers lacked one or more of the three core protection mechanisms; anti-virus software, a properly configured firewall and anti-spyware/adware software. And although 96% had heard the term spyware and 62% had anti-spyware/adware software installed, there was still 61% infected computers when scanned afterwards.

This was actually an optimistic development compared with 2004. The Online Safety Study from AOL/NCSA did in 2004[1] find 80% of US computers infected and on average 93 different spyware components on each computer. But then findings in a worldwide survey done by Webroot Software Inc. in second quarter of 2006[57], found that once again the level of infections was back on the levels from 2004. Close to 90% of checked computers now contained forms of spyware.

But we should keep in mind that e.g. the Webroot survey[57], is conducted by a company that provides security solutions against malicious software such as spyware. There

is perhaps reason to question their objectivity in the reports on spyware infections.

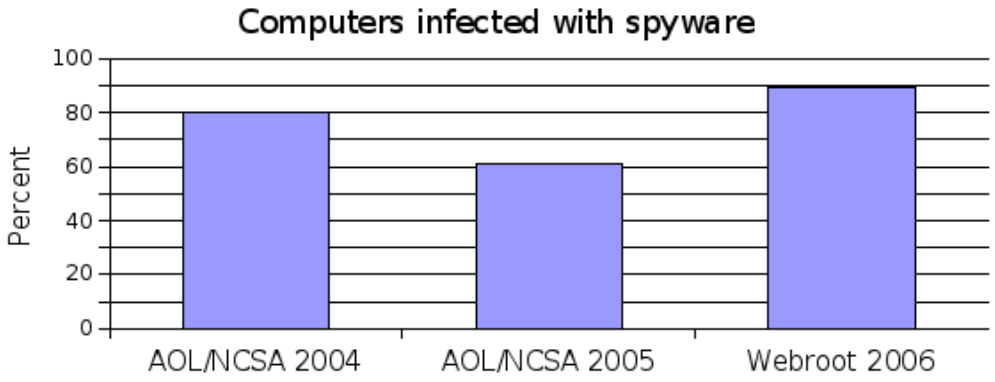


Figure 1: Spyware infections as reported by recent surveys[2, 1, 57]

The issue of spyware is by now quite well known, and it seemed that for a while the infections of spyware was decreasing. This could for example be because the adoption of anti-spyware software. But lately it has increased again, indicating that perhaps users are not vigilant enough with updates or have not chosen adequate software for the job. Also, the spyware itself might be getting more advanced and more difficult to stop.

2.3.8 Summary

Roger Thompson[55] argues that “theft through spyware could be the most important and least understood espionage tactic today”. And when we look at the number of vulnerabilities discovered last year and the latest rise in the number of infected PC’s, this seems more plausible than desired.

There exists several types of threats today and they are all becoming more and more advanced. As malware become more advanced, the security companies has to make their anti-malware products more advanced. This is an seemingly ever-lasting battle between evil and good interests. But this leaves the users as the losers, as they have no chance of keeping up to date with the emerging threats.

From the review of threats to end-users and the methods these threats use to infect computers, we have seen that most types of attacks rely on some form of unwise decision from the users. Either by not updating software, by downloading and installing unknown software, by clicking unknown links and attachments or by surfing suspect websites. When seeing how successful malware is, it is reason to think that users do still make these unwise decisions.

So perhaps the key to preventing malware in the future, is to get users to protect themselves and to make wiser decisions in from of the computer. A concluding remark from Thomas F. Stafford[49]: “There is no free lunch, and free software is just as illusory.”

2.4 Threat awareness

In the previous section, we described why user awareness is key to preventing malware. So how much do users know and are they willing to take the necessary actions to protect themselves online? We will look at a number of surveys, unfortunately mostly from the US, that have investigated what users know and think of malware threats.

2.4.1 Knowledge

The surveys done by AOL/NCSA[2, 1], asked users about several of the terms used in this paper. The term spyware is very well known, as 91% in 2004 and 96% in 2005 had heard the term. But when the users where shown a list of what spyware were found on their computers, 90% did not know what the programs were and what they did. Poston et. al[40] did a survey in collaboration with AOL, among 1006 AOL users in 2005. They found that users reported awareness of different threats; viruses(89%), spam(86%), spyware(75%), trojans(55%), worms(39%) and phishing(17%). Schmidt and Arnett[47] questioned 150 upper-division college students in 2005. They found 94% having known about the spyware threat for a year and 63% for more than 2 years.

Zhang [60] did in 2005 a survey among business majors in USA, and concluded that although most have many years of experience in using computers and the Internet, they know little about how to prevent the threats to their own privacy from malicious software. “Most users know spyware is ”out there”, but are woefully lost when it comes to preventing it or removing it.”

So it seems that the terms are becoming familiar, but users are still not very knowledgeable on the workings of spyware. To prevent spyware, people must understand how spyware operates and how they infect computers.

2.4.2 Adoption of preventive technologies

So what will it take for people to act against spyware? First we investigate what people dislike about spyware and then we look at what factors affect whether or not users will adopt anti-spyware and protect themselves.

Awad and Fitzgerald[3] identified the four most offensive deceptive behaviors of spyware; that spyware change settings on computer, that it is drive-by downloading, that it is bundled with other software and that it is slowing down computer and causing crashes. Freeman and Urbaczewski[14], did a survey including 75 undergraduate students from a US university and a Finnish business school. They found that users think that both reduced privacy and performance are important issues. They were also more concerned about privacy than performance.

Poston et. al[40] found that users are generally aware of spyware. But they are not motivated to take action or to pay for protection. Only 12% said they would subscribe to a anti-spyware service from AOL, should it be available. The users were then divided and one group were asked if they would subscribe if the service would involve a fee. Only 9% said they would definitely subscribe. The other were asked if they would subscribe if the service was free and 69% said they would definitely subscribe.

So how to get people to act against spyware? Hu and Dinev[21] found 4 factors that are

key to whether or not a user takes action against spyware. These are:

- Awareness of spyware
- Perceived usefulness of taking action
- Perceived controllability of the action
- Perceived ease of taking action

Awareness of spyware is by Hu and Dinev recognized as the most important factor, as one needs to acknowledge a problem before being able to deal with it altogether. Also, awareness was the only factor to directly influence the behavioral intention towards the adoption of preventive technologies.

Lee and Kozar[28] presents six factors in three categories that affect the adoption of an anti-spyware system. They are:

Attitude factors. The first factor is relative advantage, meaning the degree the user think anti-spyware would enhance task performance. The second factor is moral compatibility, meaning the degree adoption of anti-spyware is compatible with one's moral perception.

Social influence factors. The first social factor is visibility and is to what degree an individual sees the adoption of anti-spyware by others. The second factor is image, meaning the degree adoption of anti-spyware enhances one's image as a technical and moral leader among others.

Behavioral control factors. First factor is computing capacity, meaning the degree anti-spyware fits with one's computer and network capacity. Second behavioral factor is trialability; the degree of being able to try the anti-spyware before adoption.

Interestingly, ease of use and perceived cost was not found a significant factor in the study of Lee and Kozar. This is contradictory to the findings of Poston et. al[40] and Hu and Dinev[21] in their respective studies.

In another article, Dinev and Hu[7] did further investigations into the importance of awareness in the environment of voluntary adoption of preventive technologies and found that awareness was a strong predictor of behavioral intention towards use of these technologies. They argued that the level of technological awareness will a be key factor in making people fight spyware and other computer threats.

Their statistical analysis confirmed their argument and awareness became the central determinant of user attitude and intention to act against spyware. They findings indicate that awareness should be the at the center of information security policies and thus also in the work of getting the general public to fight the spyware problem.

2.4.3 Summary

So it seems that, at least in the US, most know of several of these threats today. And if Hu and Dinev[21] are right, we should be on our way to getting people to react to the spyware problem. Because as of yet, most do not want to make the effort of protecting themselves. It seems people will need a bigger incentive for protecting themselves.

Perhaps the examples of trojans used for Internet banking fraud, will help adoption and maintaining preventive technologies.

2.5 Chapter conclusions

A substantial amount of work has been done on studying the privacy awareness of the Norwegian population. The conclusions from these studies have not been uplifting. It seems that although people say privacy protection is important and that they claim to know of the legislation and their rights, the majority are not willing to do something to actively protect themselves or be vigilant about their own privacy. This is reflected in for instance the small number of people that exercise their rights regarding privacy and the reasons they give for not doing so.

Roger Thompson[55] argues that “theft through spyware could be the most important and least understood espionage tactic today”. And when we look at the number of vulnerabilities discovered last year and the latest rise in the number of infected PC’s, this seems more plausible than desired.

From the review of threats to end-users and the methods these threats use to infect computers, we have seen that most types of attacks rely on some form of unwise decision from the user. Either by not updating software, by downloading and installing unknown software, by clicking unknown links and attachments or by surfing suspect websites. When seeing how successful malware is, it is reason to think that users do still make these unwise decisions.

From the surveys on threat awareness in the US[2, 1, 40, 47, 60], most know of several of the threats today. And if Hu and Dinev[21] are right, we should be on our way to getting people to react to the spyware problem. Because as of yet, most do not want to make the effort of protecting themselves. It seems people will need a bigger incentive for protecting themselves. Perhaps the examples of trojans used for Internet banking fraud, will help the adoption and maintaining of preventive technologies.

3 Research method

This chapter will describe how we have approached the work of finding answers to the research questions in the introductory chapter. First we will present the overall strategy and then dig into the different parts.

3.1 Research strategy

Our research strategy consists of four sequential parts:

1. Work with what is already known.
2. Quantitative data gathering.
3. Statistical analysis of data.
4. Discussion of results.

From Creswell[6] we find that this a common approach in an exploratory quantitative research design. One first identify an area of interest, then work with existing literature to formulate appropriate research questions. The next step is to design the data gathering method and tool. After the gathering of data, statistical analysis is performed and results are discussed.

3.2 Literature

The study of related literature will be concentrated on the areas of privacy, privacy awareness, threats, and threat awareness. Sources to be used are the library at GUC, and several recognized article databases such as ScienceDirect, The ACM Digital Library, CiteSeer, and SpringerLink. In addition, several search engines and reports from security organisations will be used in the study. From the study of previous work, we found our topics for the survey.

3.3 The survey

We chose an Internet survey as our data gathering method, as we wanted to reach a large number of people in a short period of time. The following section gives a short review of the plan for the gathering of data.

The selection

The population for our survey is those of the Norwegian population above the age of 18, and are users of Internet. We have defined Internet users as those that would use Internet an average day of 2006[50]. But because of the resource limits of the thesis, we will not have the opportunity to get access to or to create a representative selection of Norwegian Internet users. Instead, we have decided on a convenience selection to be able to conduct the survey within timelimits. The requests for participants were distributed as described below.

We will send the request for participants to all 452 recruitment municipalities. The recruitment e-mails will be sent to the general e-mail address for all municipalities, with

an request to the mailstaff asking them to forward to 10 random employees of the respective government office. This is a possible total of 4520 people receiving this e-mail.

We see this as having a good chance of reaching respondents, as this e-mail address is less likely to have a strict spam-filter and is more likely to be monitored by a human. Also, we do not have the resources to gather the e-mail address of 10 random people in each municipality. But there are also disadvantages, such as policies not to forward e-mail that has no specified address or person it is going to.

The next group of respondents will be the students and employees of GUC. We will use 2 group e-mail addresses, one address that reaches all students at GUC and one that reaches all employees. This is a simple way of distributing requests, but there has been many cases of misuse of these addresses. We expect many to see our request as spam. Approximately 2000 students and employees at the school will receive the request.

The final major group of respondents will be the friends and family group. We asked friends and family to spread the e-mail to others, at work and so forth. Also, the other participants of the PETweb project were asked to forward the request for participants in their respective organisations. We estimate around 500 people will receive the request in this group.

In total, we estimate about 7000 people is the number of people possibly receiving our request to participate in our survey.

Recruitment and gathering of data

The survey will be published with online survey software from QuestionPro[52]. This approach was chosen because of the ease of use and low construction-time, compared with the construction of our own Internet form. We will be able to generate excel reports from the gathered responses and this is essential if we are to have enough time for the statistical analysis.

As mentioned, e-mail will be our primary method of distributing requests for participants. We will make a homepage for the survey, with some information and then a link to the survey at the bottom, see Figure 2 for design. This page will also include a link to the project this thesis is a part of, for those who are interested.

We see e-mail as our best approach to distributing the requests for respondents to the survey. This because of the large number of requests we need to send and because of limited resources. If we for example were to call every single municipality, we would need a week and not the day we estimate spending on sending the approximately 500 e-mails. In addition, the forwarding of an e-mail is rather uncomplicated and we think this is a major factor of success in getting people to help us. This is why we also kept focus on the fact that the participation would only take approximately 10 minutes.

Pre-survey information

Before the respondents starts answering the questions, we ask them to have the personal computer(PC) they mainly use for personal surfing in mind. We expect this to be a PC at home, but it could also be a public PC in an Internet Cafe or a PC at work.

Of the technical terms used in the survey, the English terms are often used here in Nor-

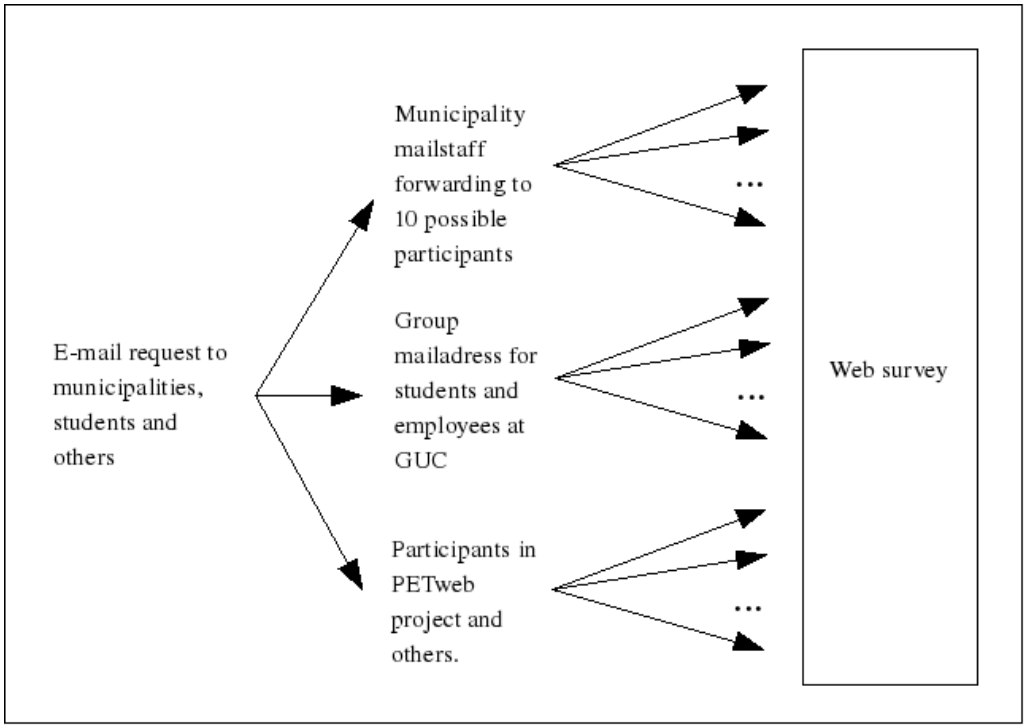


Figure 2: Recruitment process: Our design for reaching enough participants

way. Thus, we inform the respondents of instances where Norwegian terms are not commonly used or not equivalent to the English terms, we will include the English terms in parentheses.

3.3.1 The survey questions

Question theory

Haraldsen[19] describes the general thought processes that occurs when we are answering a question:

- First we interpret the meaning of the question.
- Then we bring forward the information deemed necessary for answering the question.
- We then assess what answer the information provides basis for.
- In the end an answer is formulated.

Thus, the way that the users experience of a questionnaire rests on 3 foundational aspects:

- Meaning of words and expressions used in the questions.
- What information the respondents are asked to obtain to be able to answer the questions.
- On what scale the respondents are asked to give their answer.

The fundamental rule of questions in a questionnaire is that they have to be asked in a standardized manner. This is to get as standardized and comparable answers as possible. Unstandardized questions will affect the answers in a systematic manner. For example if we formulate the various answer alternatives wrong, if the alternative in the far negative end of the scale sound more extreme than the alternative in the far positive end, we will experience a normal distribution shifted towards the positive end of the scale. This will also happen if the questions are biased.

Haraldsen[19] continues to describe different influences the respondents might be subjected to and that could affect the way the respondents answer the questions. External influences could be the questionnaire itself, like described above, or others. Haraldsen calls these other influences for general influences. This could be general opinions or beliefs about what a person should say or do in certain situations. He calls this visible or invisible audience that influence the respondent in the moment of making the choice.

So now we have a good idea of the cognitive process and influences respondents goes through and are in danger of being affected by during the survey. We then moved on to the formulation of questions and studied several other books on the subject[12, 34, 36], before compiling a list of advise:

- Ask purposeful questions and explain purpose if not easily understood. Hard questions are often interpreted to become more understandable. Do not ask for too much in terms of knowledge, memory or imagination.
- Ask concrete questions. If the question is to long and descriptive, the respondents easily overlook parts based on what they themselves think is important. And if words and terms have too wide a meaning, respondents tend almost always to interpret the meaning too narrow.
- The order of questions are important.
- Use timeperiods that are related to the importance of the question.
- Use conventional language, use complete sentences, avoid abbreviations, avoid slang and colloquial expressions and be careful of jargon and technical expressions.
- Use loaded questions only if absolutely necessary, but be careful. Do not include questions that potentially irritates, provokes, offends or hurts anyone responding.
- Avoid biased words and phrases. Keep the question balanced, it shall not lead the respondents in any direction.
- Do not ask a question that spans several topics. Do not use several information-carrying keywords. Keep It Simple Stupid.
- Avoid negative questions. Sentences that has its meaning reversed by one single word are should be avoided, as these words are easily overlooked. These types of sentences are often read as positive, so if used, the word that reverses meaning should be EMPHASIZED.

The ignoring of these basic rules leads to bias, misunderstandings and longer response times that could subsequently lead to a higher drop-out rate.

Question topics

The questions were grouped into four main topics:

1. About respondents.
2. Awareness.
3. Use of preventive technologies.
4. Interest in security measures if made available from Mypage.

The first section of questions were designed to give us the demographic knowledge needed to characterize what types of people responded to our survey. We include questions on gender, age, zip-code, education, employment, job characteristics(for those it concern) and then experience with use of computers and Internet. Thus, we have tried to compensate for the lack of a representative selection of the Norwegian Internet users.

The second section intends to measure awareness. We have used the definition from Hu and Dinev [7]; the user's following and being interested in and knowledgeable about technological issues, problems and strategies to solve them. The questions are about surfing habits, knowledge of threats and methods used by these threats, symptoms of infection on a computer and vigilance towards performing updates to preventive software and operating system. The idea behind these questions originate from the AOL/NCSA studies[2, 1] and the questions were adapted to our use and translated into Norwegian.

The third section is to investigate use of core preventive technologies, using the approximately same procedure as described in the AOL surveys[2, 1]. We adapted the core preventive technologies definition used in the AOL/NCSA study from 2005[2], where they are described as anti-virus, anti-spyware and a firewall. According to Finjan in their Web Security Trends Report from Q1-2007[11], 80% of all detected malicious code on the Internet was found in URL's categorized as advertisement. We included the popup-blocker in our definition of core preventive technologies, as popups are a large source of advertisements. We ask the respondents how secure they judge their PC to be and then ask them if they use the core security measures. Subsequent to the questions on use of measures, we include a question on the main reason they are using these measures(for those who use them) and the main reason they do not use them(for those who did not use them). These two parallel subsequent questions to each of the questions about use of the four technologies, was voluntary and in the form of an open-ended textbox. So those who want can add a comment of why they use or do not use the technology, if they have reflected on that at all.

The fourth and final section was included to see how the respondents are inclined to make use of various educational material and security measures, if provided in affiliation with the Mypage portal.

The order of topics for the survey, was chosen based on suggestions from Haraldsen[19]. The easy demographic questions was put first, to avoid loosing too many participants before they even start answering. Haraldsen argues that the first question is the most important, because most who drop out do so at the first question. Once started it is much more likely that the participant will finish the complete survey. We then continued with

the part that intends to measure awareness and then the questions about use of preventive technologies. The awareness part was put before the questions about use, because the questions on awareness require the respondents to reflect on their actions online and their level of knowledge. This is more strenuous than answering “Yes”, “No” or “Don’t know” to the use. If the awareness part is put at the end of the survey, we suspect more could give up and quit the survey. Also if asked questions about their use of preventive technologies before those on awareness, it is possible that the questions on awareness will be affected. Most people want to appear consistent in their answers[19], if someone has answered “Don’t know” to the questions on use, they may judge their knowledge lower than the actual level if this is measured afterwards.

Quality assurance

We worked closely with our supervisor and did a pilot survey before sending out the requests for participants to the final survey. When we felt that the survey was nearing completion, we sent out the survey to 10 people with various levels of expertise on the area. We requested them to take the survey and comment on anything they did not understand or thought odd in any way. The comments was then reviewed and changes were made to the survey. The pilot participants were consulted about every change regarding their comments and the survey was finalized, see Figure 3 for design.

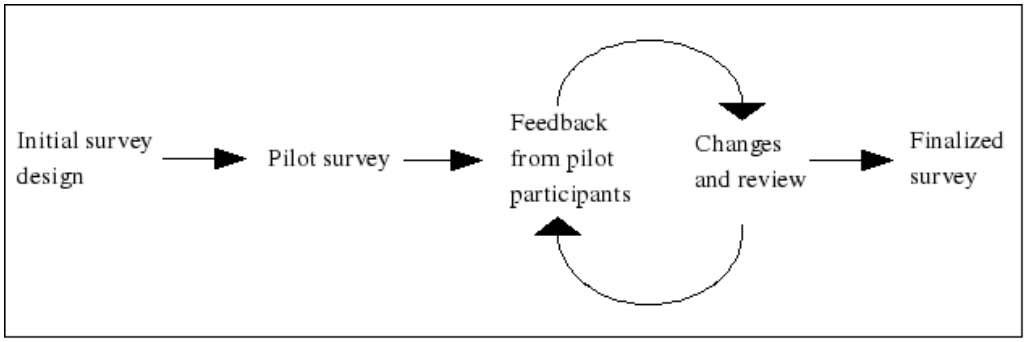


Figure 3: Quality assurance: The use of a pilot survey to avoid question bias

Design issues

We focused on keeping the completion time at about 10 minutes. This would not give us very thorough measure of awareness, but as the respondents would be recruited on a voluntary basis we focused on getting as many completed forms as possible.

We chose to use almost exclusively mandatory questions. Because we had the use of preventive technology questions in the last part of the survey, we needed completed surveys to be able to use the results. One of the major points of the data gathering was to compare use of preventive technologies at various levels of awareness. Thus, a only partially completed form would be useless. The only non-mandatory questions were the open text questions following the questions about use of technologies.

We used Norwegian as the only language option for our survey. We simply did not have the resources available for a complete translation and creating of a duplicate survey. We did not have the time needed for this and the student edition of the QuestionPro

software[52] allowed for only one active survey at a time. Unfortunately this excluded any non-Norwegian speaking individuals from participating in our survey. But this number was estimated as so small, that we chose to limit ourselves to only one language.

3.4 Statistical analysis of data

This section gives a short summary of the strategy and statistics we will use in the statistical analysis of the data from the survey.

Strategy

We plan for the following strategy in the analysis of the data from the survey:

Survey response. This part will include available information on number of completed forms, drop-out rates and completion time.

Descriptive analysis. Interesting variables will be described with means, standard deviations, range of scores, and similar.

Our selection. We will use the demographic questions to compare our selection with the population.

Awareness. Awareness will be computed and differences in selection subsets will be investigated. We will also test for normality of our data, although we assume here to get normally distributed scores. Factor analysis and reliability tests for internal validity will be performed.

Use of preventive technologies. Differences in selection subsets will be investigated.

Compare awareness with use. Comparison of means and ANOVA analysis will be used to determine connections between awareness and use. Correlation and regression analysis will be used to explain the effect of awareness on use.

Interest in measures from Mypage We will describe the interest in prospective educational material, software and security services from Mypage.

Tools

We plan to base our analysis on the tools available in the MS Office Excel equivalent Openoffice Calc. But this depends on how much response we will get on our survey. Should we approach 1000 responses we expect to be considering a different tool as the spreadsheets would get quite large and complex. We do not have great experience with working large datasets in Excel or Calc, so we were prepared to move on to a more powerful statistical tool such as SPSS[48] or PSPP[13].

3.5 Discussion

In the discussion chapter we will discuss the results from the analysis in view of our research questions. We will also look back on the methods chosen here and comment on how well they performed for our use.

4 Survey response

Respondents	
Viewed welcome page	1086
Started answering	936
Completed	784
Completion rate	83.76%
Drop Outs (After starting)	152
Average time taken to complete	9 minutes

Table 1: Survey statistics

"Viewed first page" are those who only viewed the welcome page of the survey and then closed the survey window. Those in the started answering group began answering questions, but did not finish. Those in the completed group finished all required questions.

We are quite pleased with the number of respondents we got in our survey. The total of 784 completed survey forms give us a good source of data for the statistical analysis we will be conducting in the next chapter. According to [29], a sample size of 400 is adequate for any population size above approximately 5000 units, indicating our 784 completed forms should be sufficient.

4.1 Respondents

The goal of any sampling from a population is to get an as representative selection of the population as possible. As our survey was not distributed to a representative sample of the population of Norwegian Internet users, we made a few comparisons between our sample and the Norwegian Internet users.

Gender

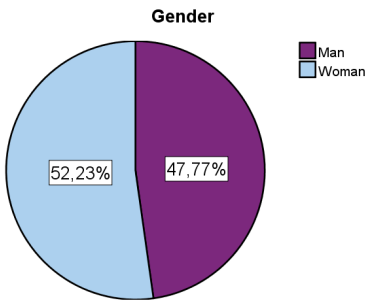


Figure 4: Gender distribution for our respondents

We have an approximately even distribution between male and female respondents, but is somewhat skewed compared with the internet user gender distribution of 55% men and 45% women[50], see Figure 4 for our gender distribution. But if we look at the

gender distribution for users of Internet banking, online shopping and similar[50], we find this to be 48.4% for men and 51.6% for women.

Age

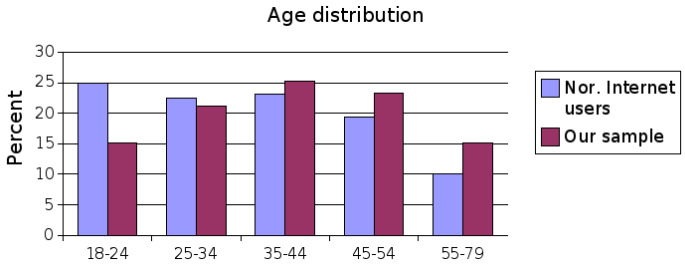


Figure 5: Age distribution for our sample and the Norwegian Internet users

The age distribution is also similar to the age distribution of Norwegian Internet users aged 18 and above, Figure 5. It appears we have slightly older respondents, than the average for Norwegian Internet users. This could be explained by the fact that we have primarily distributed the survey to employees and college/university students. We included an option for those aged 17 and below in the survey, but did not get any respondents in this group. The explanation for this is probably the lack of people at this age in both in college/university and full-time employment. There is no significant difference in age for men and women in our sample.

Education

When comparing the education level of our respondents to that of the Norwegian Internet users[50], we see that our respondents have a higher degree of education than average. One reason for this could be the higher age average for our sample.

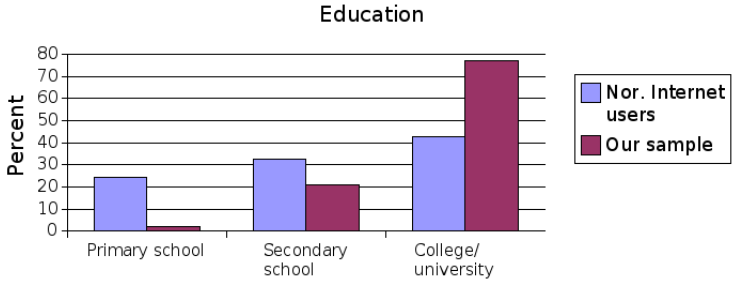


Figure 6: Education distribution for our sample and the Norwegian Internet users

Employment

No surprises here, we have a majority of governmental employees and students which corresponds well to the way we distributed the requests for participants.

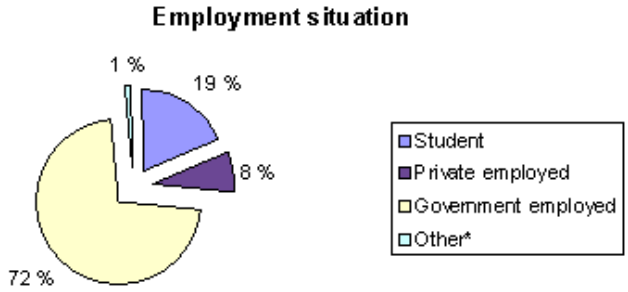


Figure 7: Employment distribution for our respondents

*The "Other" category consists of those who stated they were self-employed, retired, unemployed, homemaker, compulsory military service or social welfare.

Computer and Internet experience

Apparently we have rather experienced respondents, nearly 78% has 11 or more years of experience with using a PC, see Figure 8 and nearly 83% has 6 or more years of experience with surfing the Internet, see Figure 9. Unfortunately, Statistics Norway did not have any similar numbers from the Norwegian Internet users for us to compare with.

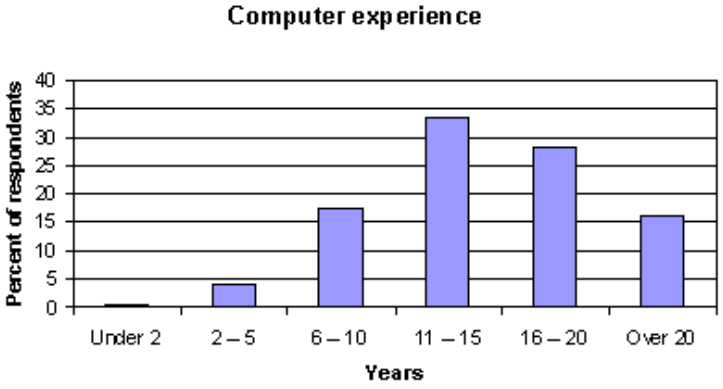


Figure 8: Computer experience distribution for our respondents

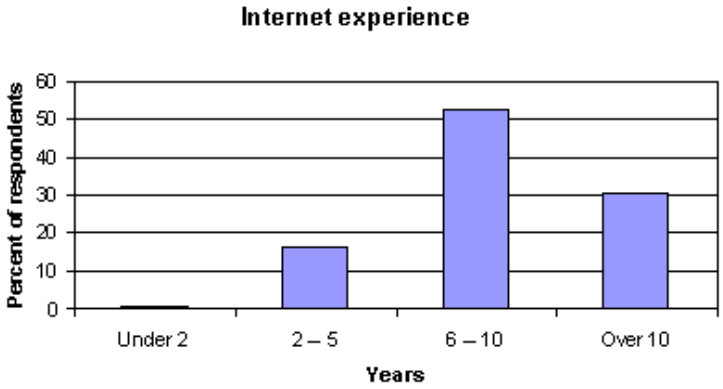


Figure 9: Internet experience distribution for our respondents

Geographical distribution

County	Frequency	Percent of total
Østfold	43	5.76
Akershus	47	6.29
Oslo	25	3.35
Hedmark	40	5.35
Oppland	151	20.21
Buskerud	46	6.16
Vestfold	23	3.08
Telemark	26	3.48
Aust-Agder	15	2.01
Vest-Agder	14	1.87
Rogaland	32	4.28
Hordaland	49	6.56
Sogn og Fjordane	29	3.88
Møre og Romsdal	58	7.76
Sør-Trøndelag	22	2.95
Nord-Trøndelag	29	3.88
Nordland	43	5.76
Troms	24	3.21
Finnmark	31	4.15
Total	**747**	**100**

Table 2: Geographical distribution

Based on what zip codes the respondents gave during the first part of the questionnaire, we have made an table of the geographical distribution of our respondents, see Table 2 (Also see Appendix C for illustration). There were a total of 747 valid zip codes registered, as we did not have a way of checking that a valid zip code was entered. The missing values are possibly due to mistyped numbers or similar. The rather large group of respondents from the Oppland county is explained by the fact that GUC is situated in this county and students/employees at GUC was a large part of the selection. Other than that, we have got respondents from all counties.

5 Statistical analysis

5.1 Preparation

The QuestionPro[52] online survey software would export the survey results to a Excel document, including descriptive statistics for the survey and each question, and the raw data. Seeing the number of respondents we got for our sample it was obvious that we would experience difficulties with the excel-clone Calc for processing. The spreadsheets would become uncomfortably large and complex quickly, so we decided to move to the SPSS software[48]. We used books like [15, 8] for the introduction to statistics with SPSS.

As discussed in subsection 3.3.1, only a few of the questions were not mandatory and as we see from the completion rate in Chapter 4 this worked out satisfactory with only a 16.24% drop-out rate after beginning answering questions. SPSS handles cases of missing values(incomplete survey forms in our case), so we did not remove incomplete survey forms before converting the data to a SPSS data file.

In the raw data the answers to the questions were represented as a value between 1 and *n*, *n* being the number of answer alternatives. When we got the raw data, we realized that a few of the question alternatives needed a value recode. For most of the questions we chose answer alternatives ranging from a "negative" or "least aware" choice(Example from knowledge on security measures: "I have never heard of it") to a "positive" or "most aware" choice(Example from knowledge on security measures: "I can both install and configure"). This ordering of alternatives makes a less aware choice score lower and a more aware choice score higher.

Example 1: Question on "Bad" habit:
"Do you click advertisement banners or ads in popups?"
Answer alternatives:
1: Often
2: Now and again
3: Rarely
4: Never

Example 2: Question on "Good" habit:
"Do you read EULA before installing software?"
Answer alternatives:
1: Always
2: Often
3: Rarely
4: Never

From the examples above we see that the question in example 2 will be given a high score for the least aware choice and a low score for the most aware choice, so we recoded the values for the questions affected by this; questions 11, 12, 26, 27, 29 and 30, see

Appendix B for complete survey form. What was done specially for question 29 and 30 was the recoding of both the “Yes” and “No” answer to give 2 points and the “Don’t know” answer to give 1 point. The reason for this was that we only wanted to know whether or not the respondents had noticed this or not, see Appendix D for full recode table.

5.2 Awareness

When we had prepared the raw data, we intended to range the awareness for each respondent as described in Section 3.4, with the inclusion of question 9-27, 29, 30, 32-35 and 37 in the awareness score. But we first wanted to do some tests on our awareness score.

5.2.1 Factor analysis

To test if our questions measured awareness or at least the same construct, we did a factoranalysis of the questions included in the awareness score, see Table 3 for variable numbers used in subsequent discussion. As seen in Figure 10, we got one factor having an eigenvalue of 10.13 and explaining almost 39% of the variance in the awareness score. We also got 6 more factors with eigenvalues above 1, which is the default setting for accepting factors in SPSS[15]. The eigenvalue is a measure of the standardized variance for the component, compared to that of the original variables. Since we have 26 variables(questions), we have a total standardized variance of 26. An component eigenvalue lower than 1, indicates that a factor explains less of the variance in the construct, than any original variable.

Question to variable conversion													
Question no.	9	10	11	12	13	14	15	16	17	18	19	20	21
Variable no.	1	2	3	4	5	6	7	8	9	10	11	12	13
Question no.	22	23	24	25	26	27	29	30	32	33	34	35	37
Variable no.	14	15	16	17	18	19	20	21	22	23	24	25	26

Table 3: The question to variable recode used for the factor analysis

But when we looked at the load values(this roughly the correlation between a variable and a component) for the different variables(each variable representing one question from the survey) in each component, we decided to keep only the first component for our further analysis, see Appendix E for complete factor analysis matrix. For factor 2 through 7, there were very few good component loads for the variables. But we found some interesting results from the factor component matrix. In the first component the variables 1, 2, 3, 4, 20, and 21 loaded rather low, between -0.082 and 0.333. All other questions got component loads above 0.5, which is the level we have set for including a variable in a component. But then when looking at component 3, 4, and 7, we find variable 1 and 2 scoring above 0.5 on component 4, variable 3 and 4 scoring above 0.5 on component 3, and variable 20 and 21 scores above 0.5 on component 7. No other variables loaded above 0.5 on either of three additional components.

From these result there is reason to suspect that component 1 is measuring a construct mainly based on knowledge and that the three additional components represent other constructs. The variables 1, 2, 3, 4, 20, and 21 intend to measure certain habits and attentiveness to change. Perhaps these variables are also affected by the knowledge variables,

but we will not look further into this in the current report.

Eigenvalues

Component	Initial Eigenvalues		
	Total	% of Variance	Cumulative %
1	10,132	38,970	38,970
2	1,885	7,251	46,222
3	1,680	6,460	52,682
4	1,327	5,102	57,784
5	1,298	4,991	62,775
6	1,075	4,134	66,909
7	1,043	4,010	70,918

Extraction Method: Principal Component Analysis.

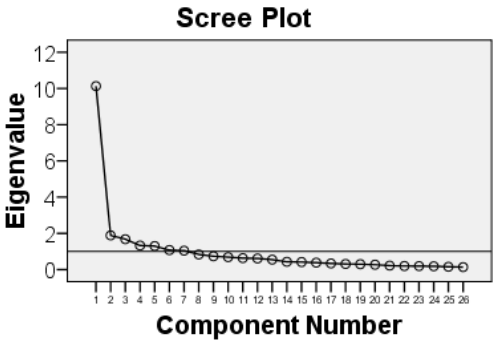


Figure 10: Eigenvalues and screeplot from factor analysis

To get the best results in the subsequent correlation and regression analysis we decided to keep only the variables that loads above 0.5 on one factor and at the same time not above 0.3 on another. This type of selection is done to remove unclear variables from the measurement construct, in our case the awareness. The levels are suggested by [15] and by doing a preliminary regression analysis, we found these levels to give us better results. We have an example of this process in Figure 11.

Partial Component Matrix

	1	2
ClickAds	,156	,036
InstallPrograms	,148	,042
ReadEULA	-,082	,070
UnderstandEULA	,121	,082
KnowVirus	,635	,380
KnowTrojan	,734	,127
KnowSpyware	,784	,139
KnowKeylogger	,768	-,410

Figure 11: Partial matrix from the component extraction in the factor analysis

The ClickAds variable is not loading above the level of 0.5 in the first component and is thus rejected. The KnowVirus variable is loading high enough on the first component, but is also loading 0.380 on another and is rejected. KnowSpyware has a good load on the first component and no load above 0.3 on any other and is accepted. KnowKeylogger is loading above 0.5 on the first component, but is loading -0.410 on the second component and is rejected. Whether the second load is above 0.3 or below -0.3 is not of importance, we look for any correlation between the variable and another component and not in what direction they are correlating.

From this filtering process we ended up with the variables 6, 7, 13, 14, 16, 23, 24, 25 and 26. Before the variable reduction, we had a Cronbachs alpha value of 0.923 and after the reduction we had a value of 0.908. This value is designed to measure internal consistency on a scale from 0 to 1, and a value of 0.7 or higher is considered a good

result[15].

5.2.2 Normal distribution

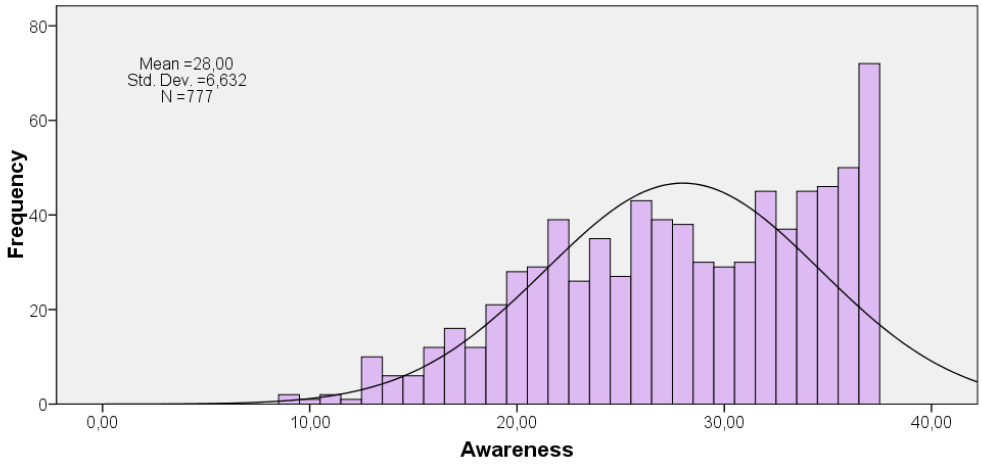


Figure 12: Awareness score for our sample

When the scores where grouped and presented in an frequency table we saw that the overall score looked pretty normal distributed for our sample, see Figure 12. The values for skewness was -0,413 and the value for kurtosis was -0,734, and according to [15] a value between +1 and -1 is acceptable.

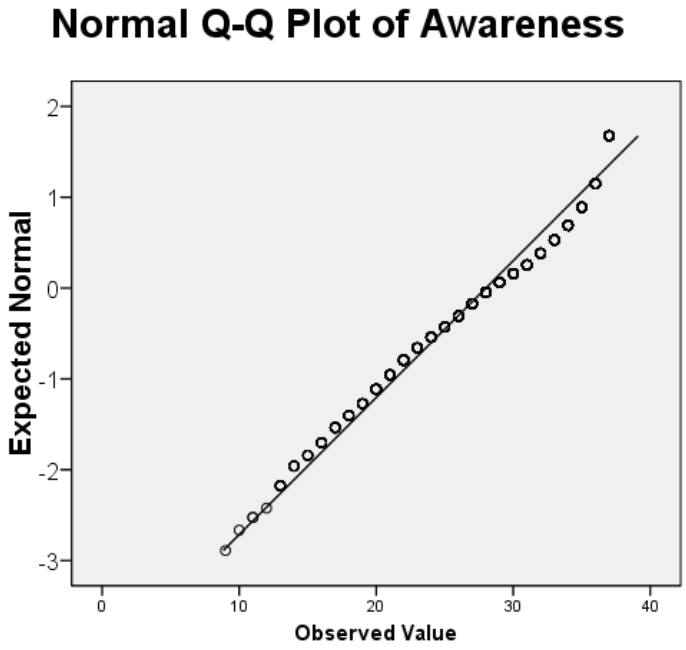


Figure 13: Normality plot for awareness score

Using the normality plot, see Figure 13 we determine the sample to normal distributed.

This is confirmed by the normality tests of Kolmogorov-Smirnov and Shapiro-Wilk, see Figure 14. Both tests scoring higher than 0,05 and thus no significant deviation from normality.

Tests of Normality

	Kolmogorov-Smirnov[a]			Shapiro-Wilk		
	Statistic	df	Sig.	Statistic	df	Sig.
Awareness	,107	777	,000	,950	777	,000

a. Lilliefors Significance Correction

Figure 14: Normality tests with Kolmogorov-Smirnov and Shapiro-Wilk tests

We observed a mean of 28 points on the scale from 9 to 34 points. But this overall score is not that easy to put into perspective and because it includes scores from so many different types of questions, we cannot really say that a specific level between 9 and 34 is good enough when looking at the awareness score on its own. We will not explain the awareness value further in this section, because we argue that the awareness level needs to be viewed in context with the use of preventive technologies. In Section 5.4 we will look for connections between awareness and use. The following section will look into differences in sample subsets.

5.2.3 Differences for sample subsets

As we saw in Chapter 4, we have a somewhat skewed sample of the Norwegian Internet users. Our sample differed from the population in both age and education. But we still wanted to look for differences in sample subsets, that might help us better predict how the population of Norwegian Internet users will score.

Difference in gender

When comparing the awareness score for the genders we saw a mean of 30.3 points for men and 25.8 for women, which is significantly different at the 5% level, see Figure 15.

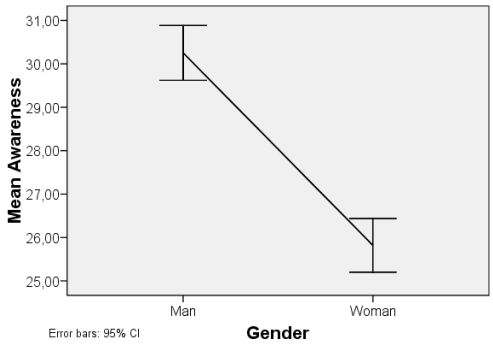


Figure 15: Mean awareness score by gender

But we see the same differences in gender when it comes to experience with both PC and Internet. We got significant higher experience for the male part of our sample. Further investigations into experience affecting on awareness, showed the results in Figure 16. A significant linear increase in awareness as experience with PC and Internet increases.

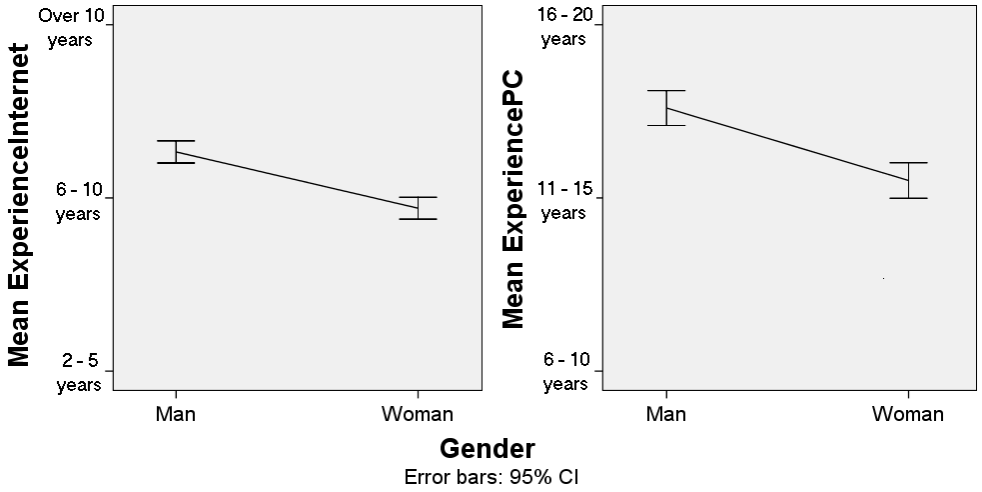


Figure 16: Experience with Internet and PC by gender

Difference in age

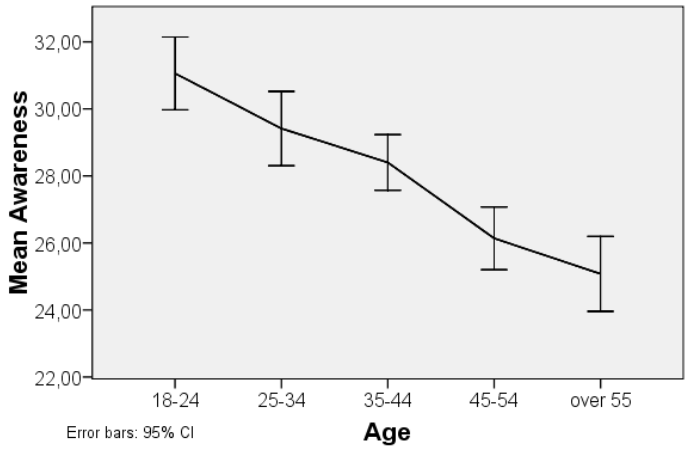


Figure 17: Awareness score by age

We found a significant linear decrease in awareness score as age increases in our selection.

Difference in education

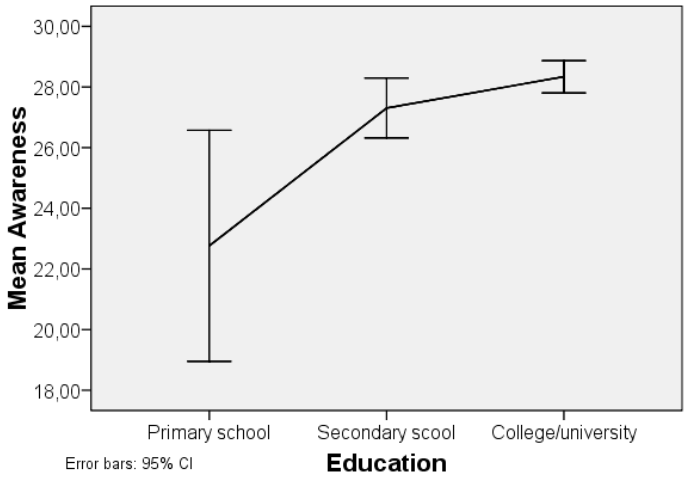


Figure 18: Awareness score by education

We also found a linear significant increase in awareness as education increases.

Difference in experience

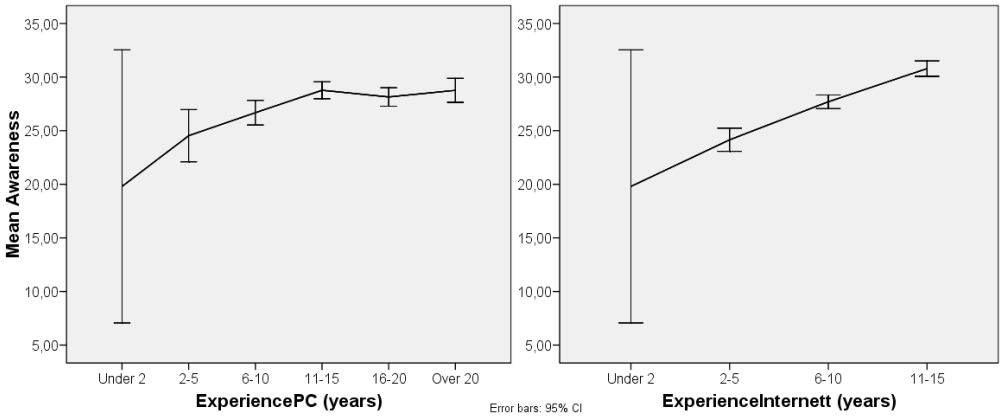


Figure 19: Awareness score by Internet and PC experience

We see another linear significant increase in awareness as both experience with PCs and Internet increases.

5.3 Use of preventive technologies

The first look at what the respondents answered on the questions on what preventive technologies the make use of, we get the results shown in Table 5.3.

Table 4: Descriptives on the different preventive technologies

Use of preventive technologies		
	Frequency	**Percent**
Anti-virus		
User	767	94.1
Non-user	38	4.7
Don't know	10	1.2
Anti-spyware		
User	420	52.1
Non-user	128	15.9
Don't know	258	32
Firewall		
User	578	72.3
Non-user	127	15.9
Don't know	95	11.9
Popup- blocker		
User	529	66.5
Non-user	125	15.7
Don't know	142	17.8

We see that use of anti-virus is very good, close to 95% is a very good result. The the use decreases with 71% for a firewall, 67% for popup- blocker and 52% for anti-spyware.

5.4 Connections between awareness and use

5.4.1 Mean awareness comparison

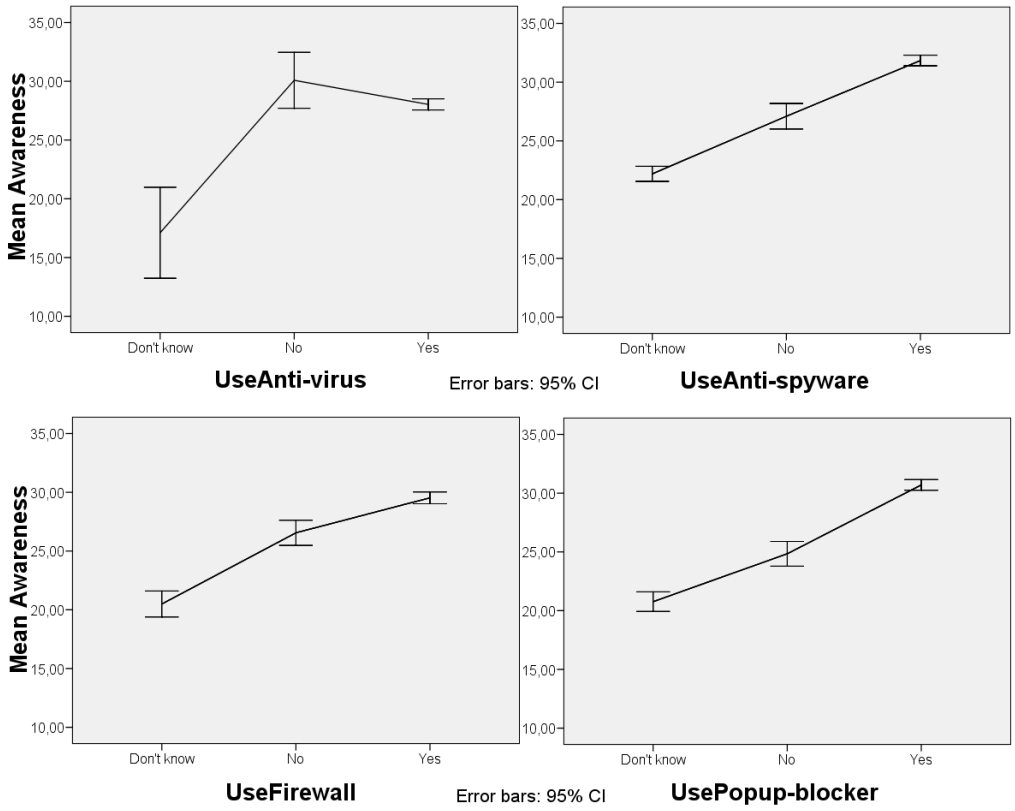


Figure 20: Comparison of mean awareness score sorted on use of technologies

First we looked at the difference in awareness mean for the unaware, non-users and users of the four preventive technologies, see Figure 20. The ANOVA analysis showed significant differences between the three groups of respondents for all four preventive technologies. For all technologies, there was a linear significant increase from the “Don’t know” group to the “No” and from the “No” group to the “Yes” group. The only technology that did not have a 1% significance level was anti-virus.

5.4.2 How well does awareness explain use?

We are interested in exploring the connections between awareness and use of preventive technologies. Looking at the correlation values, see Table 21, between awareness and the four preventive technologies, we note low correlation on anti-virus and strong correlation between the other three.

Correlations

		UseAnti-virus	Use Anti-spyware	UseFirewall	Use Popup-blocker
Awareness	Pearson Correlation	,081*	,651**	,436**	,599**
	Sig. (2-tailed)	,024	,000	,000	,000
	N	777	777	777	777

*. Correlation is significant at the 0.05 level (2-tailed).

**. Correlation is significant at the 0.01 level (2-tailed).

Figure 21: Correlation between awareness, and the 4 preventive technologies

Looking closer at the effect of awareness on the 4 technologies, we use regression analysis and find quadratic equation performs better than the linear on for all technologies, see Table 5.

R square values for linear and quadratic equations

Equation	Anti-virus	Anti-spyware	Firewall	Popup-blocker
Linear	0.007*	0.424**	0.190**	0.359**
Quadratic	0.032**	0.425**	0.199**	0.368**

Table 5: Results from regression analysis

* indicates significant at 5% level, ** indicates significant at 1% level.

From this table we see that our awareness score, using the quadratic regression equation, explains 3.2% of the variance in use of anti-virus, 42.5% of the variance in use of anti-spyware, 19.9% of the variance in use of a firewall, and 35.8% of the variance in use of a popup-blocker.

5.4.3 Average use at different levels of awareness

We then grouped the awareness scores and calculated the average use of each technology for each level of awareness.

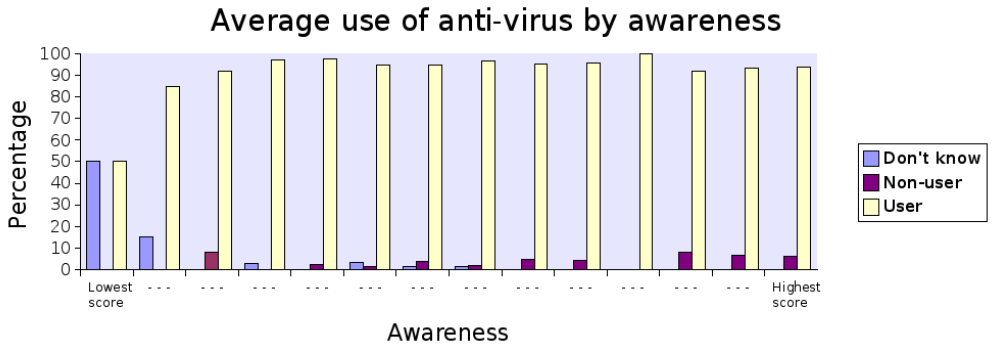


Figure 22: Average use of anti-virus by awareness score

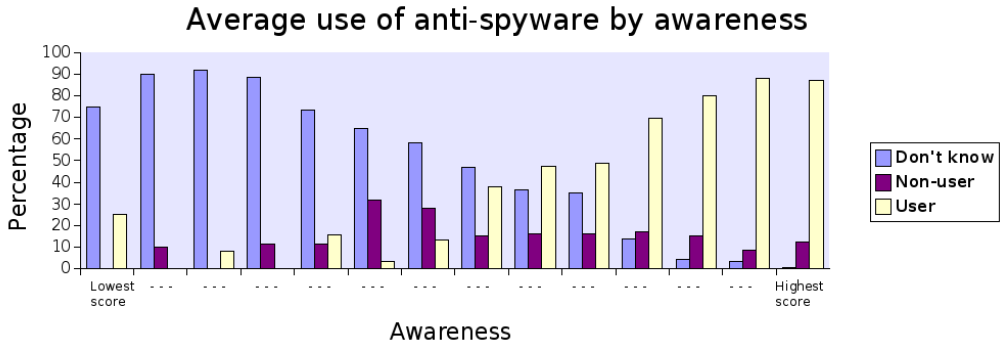


Figure 23: Average use of anti-spyware by awareness score

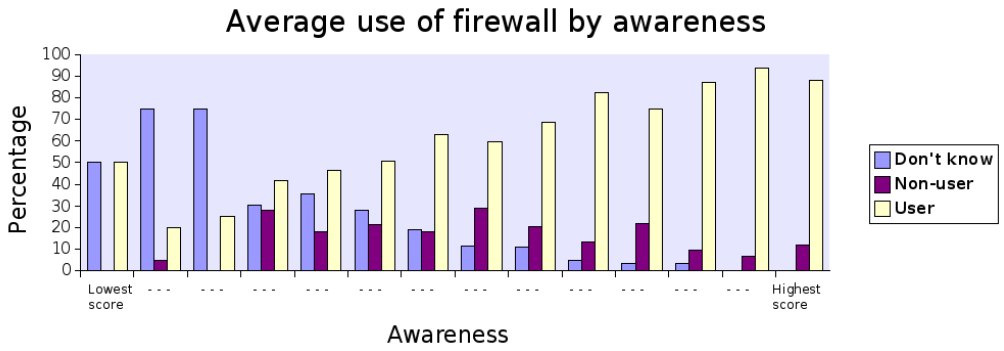


Figure 24: Average use of firewall by awareness score

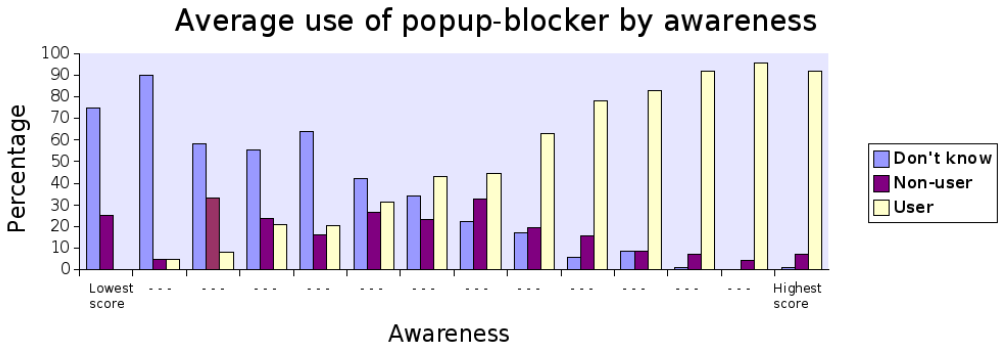


Figure 25: Average use of popup blocker by awareness score

5.4.4 Trends in use of preventive technologies

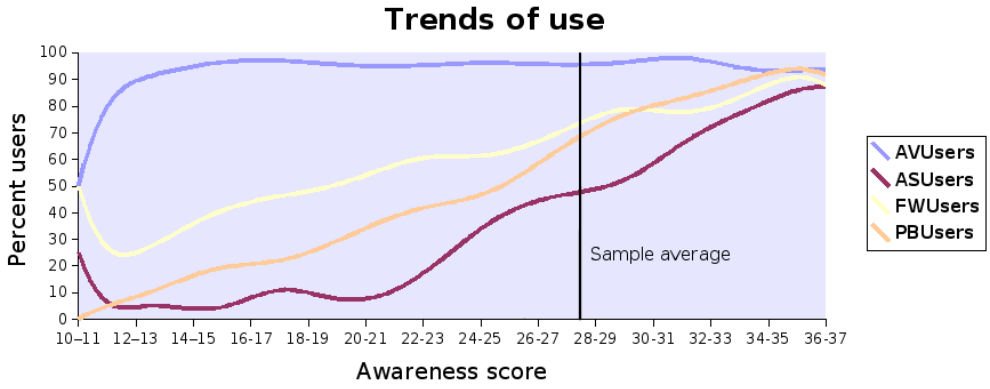


Figure 26: Trendlines for use

These trendlines show us at what level of use of anti-virus(AV), anti-spyware(AS), firewall(FW) and pop-up blocker(PB) could be expected users at the different levels of awareness. To decide what level of awareness is satisfactory, one needs to decide what level of use is desired. We have grouped the awareness score and smoothed the lines to get a more readable figure.

5.5 Interest in security measures in affiliation with Mypage

The final section of questions was included to get an idea of what kind of measures the potential users of Mypage were the most interested in, should the portal decide to make security resources available for its users.

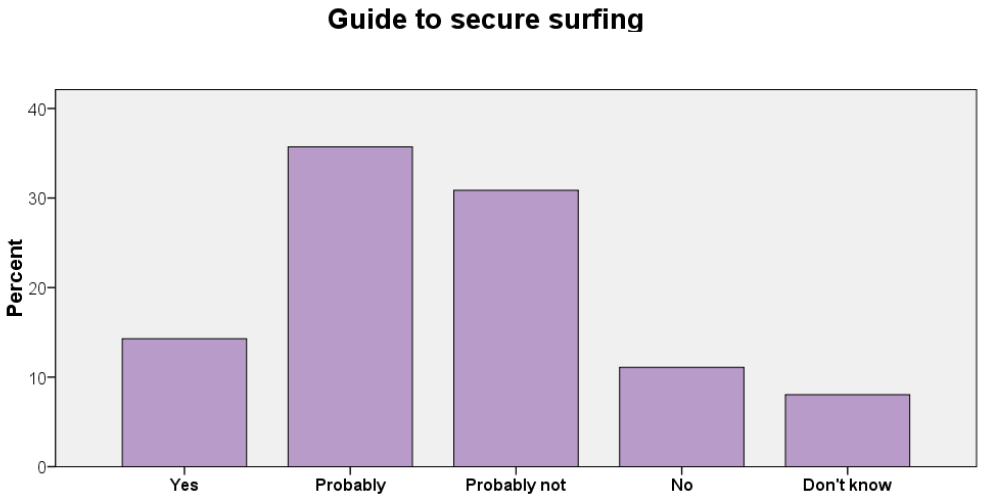


Figure 27: Interest in guides to safe surfing

The question was “Would you make use of guides to safe behaviour online?” and we found 50% positive to this service, 42% negative and 8% unsure.

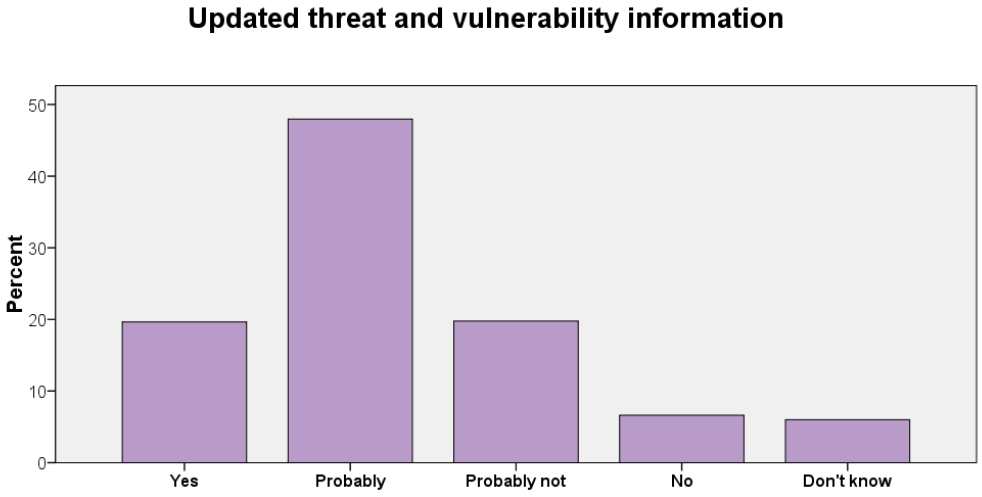


Figure 28: Interest in updated threat information

The question was “Would you make use of updated threat information?” and we found 67.6% positive to this service, 26.4% negative and 6% unsure.

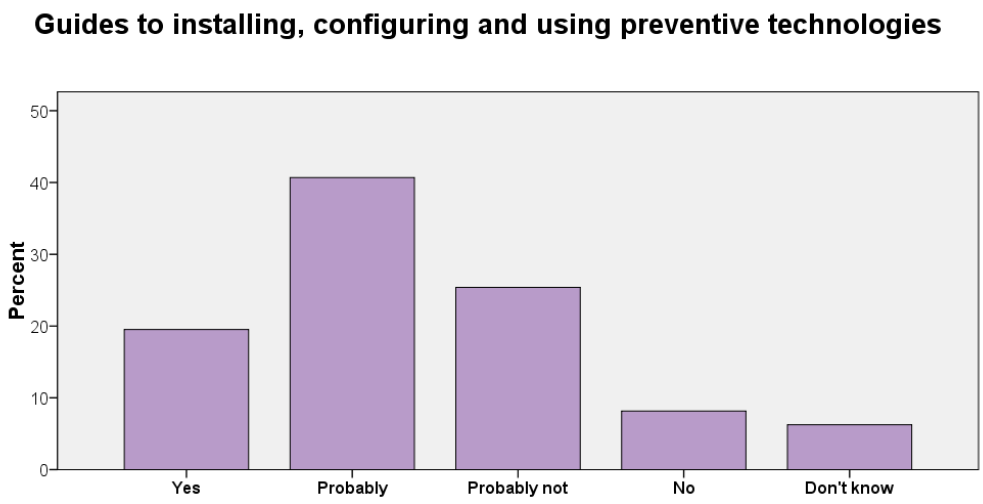


Figure 29: Interest in guides to preventive technologies

The question was “Would you make use of guides to installing, configuring and using preventive measures?” and we found 60.2% positive to this service, 33.5% negative and 6.3% unsure.

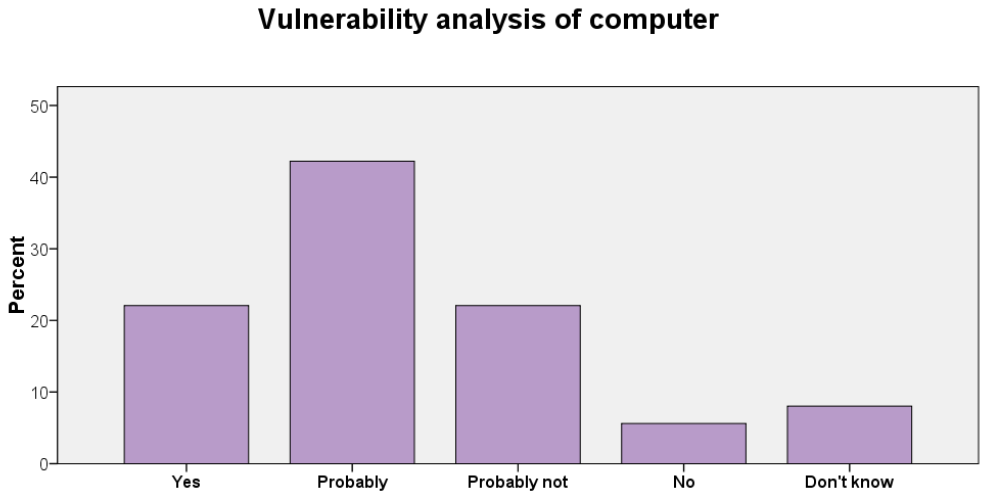


Figure 30: Interest in a vulnerability check service

The question was “Would you make use of a vulnerability-check service?” and we found 64.3% positive to this service, 27.7% negative and 8% unsure.

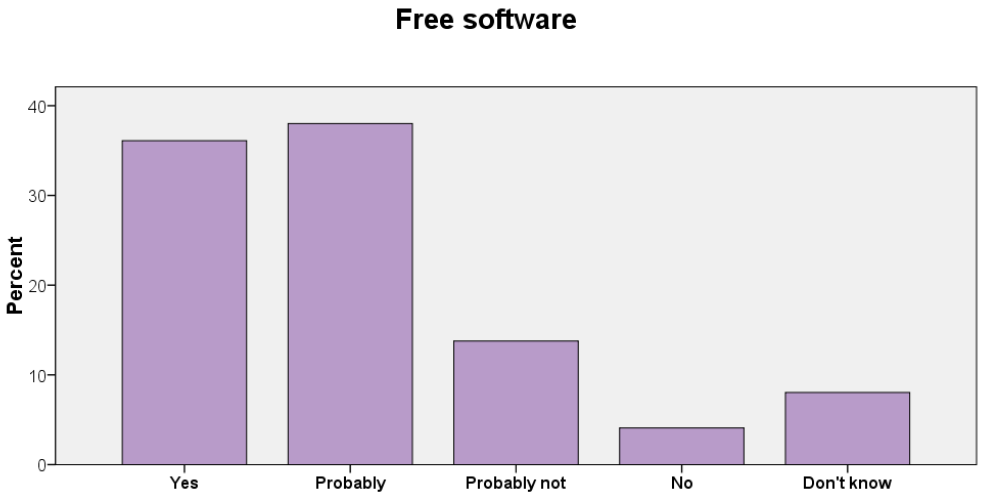


Figure 31: Interest in free preventive technologies

The question was “Would you make use of free preventive technology downloads?”and we found 74.1% positive to this service, 17.9% negative and 8% unsure.

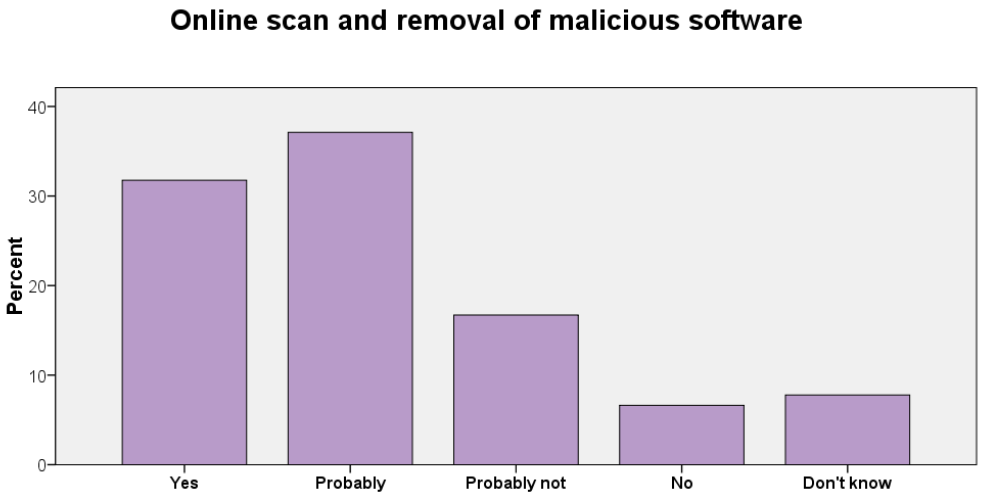


Figure 32: Interest in a online scan service

The question was “Would you make use of a online scan and removal of malicious code service?” and we found 68.9% positive to this service, 23.3% negative and 7.8% unsure.

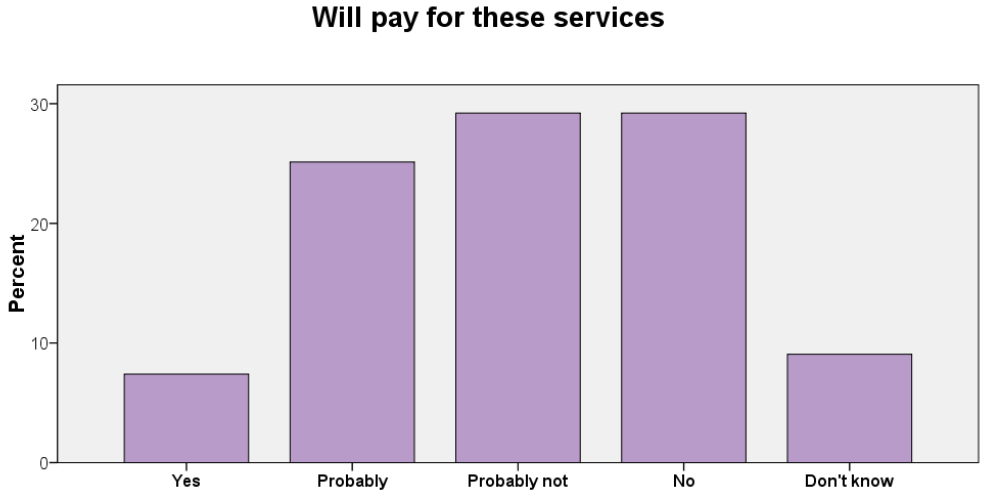


Figure 33: Willingness to pay for the services

The question was “Would you be willing to pay for these services, either as a one-time fee or as a subscription?” and we found 32.5% positive to this statement, 58.4% negative and 9.1% unsure.

6 Discussion

This chapter will discuss the results found in the previous chapter. We start off with the discussion of the differences between our sample and the population of Norwegian Internet users, found in Section 4.1. We then look at the results found in the statistical analysis and finally we discuss the research methods used in this thesis.

6.1 Sample versus population

Our sample of the Norwegian Internet users did not have the exact same characteristic as the population, as seen in Section 4.1. The gender distribution differs somewhat from the population of Norwegian Internet users, but is very similar to that of Internet banking users from 2006[50]. The age average for our sample is higher than for the Norwegian Internet users and our sample has a higher average education level, see Section 4.1.

But we looked at how much the mean awareness score differs in subsets sorted on age and education, see Subsection 5.2.3. Increase in age has a negative effect on awareness score and a higher education has an positive effect on awareness score, lower age . We thus argue that the effects of our difference from the population is somewhat neutralized. We have a larger deviance from the population in education, than in age. But in return the awareness is more affected by the age than the education. We used age and education as independent variables on the dependent variable awareness in a regression analysis. We found the quadratic equation for age explaining 8.9% of the variance in the awareness score and education explaining 1.8% of variance in awareness, indicating that age is more important for the awareness than education. Based on this we argue that this to some degree evens out the differences between our sample and population.

But still it should be made clear that the results found in the analyses mainly apply to our sample and that the instrument should be applied to a representative selection for valid results applying to the entire population.

6.2 Results from statistical analysis

6.2.1 Awareness

The intended awareness measurement in this thesis was reduced somewhat from its origin. We had 26 questions in our survey, which we planned to base our awareness score on. These were reduced to 9 after the factor analysis, due to the criteria used for component loadings. It should be noted that several of the other questions were just outside the criteria of being accepted as part of the score, see Appendix E for full factor component matrix.

Since all questions in each group was asked identically, there is a possibility that the respondents were uncertain of e.g. the types of threats asked for. We asked about knowledge on viruses, trojans, spyware, keyloggers, phising, and social engineering. Only questions on trojans and spyware were accepted in the factor analysis. We argue that these might the most commonly known threats, due to publicity in the media, see Appendix A.

A quiz form of testing knowledge on these threats would give a more correct measurement of the knowledge. Perhaps our questions do not measure knowledge sufficiently precise, since people were asked to rate their own knowledge on areas their are uncertain of. We might have been affected by the bias mentioned in [19] as people wanting to "color" their answers a bit and not admit lack of knowledge.

We will not discuss the awareness score further as it needs to be viewed in comparison with the use of preventive technologies. This because it is argued to be an underlying cause of the use and we want to keep a focus on the result of the awareness, in our case use of preventive technologies.

6.2.2 Use of core preventive technologies

To sum up the results of preventive technology use for our sample; anti-virus is used by 94.1%, firewall by 72.3% and popup-blocker by 66.5%, and anti-spyware by 52.1%, see Table 5.3.

Anti-virus is pretty much used by everyone, this is probably due to the fact that the threat of the virus has been known for a long time. From the open text question we learned that those who do not use anti-virus, made this choice based on the operating system they used. Most state using either Linux or OS X and do not regard the threat to these operating systems as great enough for the time being to install anti-virus. But some state that it is because they use a Mac computer and that there is no threats to Mac(OS X). This is not true, as malicious code developed for the OS X is out there. And although the natural approach is to develop malicious code for the largest target group, we could see more platform independent or OS X malicious code in the future.

Firewall is utilized by 72.3%. A quite high number and this is not unexpected as e.g. Windows has a built-in firewall, and also wireless routers and broadband routers often have a built-in firewall. From the open text question we see that answers like "too little knowledge", "perceived difficulty of installing" or "too much hassle with access restrictions" are repeated for those who do not use a firewall.

Popup-blocker is used by 66.5%. One could perhaps expect that this number should be higher as major Internet browsers such as Internet Explorer, Firefox, Opera and Safari now have a built-in popup-blocker. From the open text question we see that lack of knowledge and excessive restrictions on popups they need or want are main reasons why people do not use it. Perhaps the second reason people give for not using a popup-blocker might be due to lack of knowledge as well. Most popup-blockers can be configured to allow popups from certain websites, making the excessive restriction argument less valid.

Use of anti-spyware is at 52.1%. This alone is rather worrying; we quote Thompson[55]: "theft through spyware could be the most important and least understood espionage tactic today". The results from the survey done by Webroot Software Inc.[57] also adds to this worry; in the second quarter of 2006, close to 90% of checked U.S. home computers contained forms of spyware. But we have a large group of people not knowing whether or not they use anti-spyware(32%), so there might be a larger number of people being unknowingly protected by e.g. built-in anti-spyware in other security solutions.

The AOL/NCSA US study[2] from 2005, found differences between what users claimed to have of protection and what protection was actually installed on their computers. 81% of respondents claimed to have anti-virus protection, 67% claimed to have a firewall on the computer and 70% claimed to use a popup-blocker. Unfortunately it seems to be inconsistencies in the types of preventive technologies enquired about and scanned for in this survey. But the scan subsequent to the survey, revealed 83% use of anti-virus, use of firewall at 78% and 62% had anti-spyware software installed. The comparable numbers of anti-virus and firewall indicates that the actual use is higher than what people are aware of. Perhaps this is the situation in Norway as well, but only a similar scan of user computers would confirm this.

Whether or not the use of preventive technologies for our sample is satisfactory or not, depends on the actual percentage of spyware infected computers. It is a fact that if you are extremely careful about anything relating to Internet or even keep your computer disconnected from the Internet, you could be fine without any extra protection. 91.3% of our sample states to rarely or never click on popups or ads on the Internet and 97.5% rarely or never install programs recommended by these popups or ads. Only a scan of user computers can give a decisive answer to whether or not the Norwegian Internet users are well enough protected against threats from the Internet.

One uplifting result is that 82.34% in our survey state they update their security software along with the operating system regularly or everytime an update is available.

6.2.3 Awareness versus use of preventive technologies

This section will look into the results we got from the statistical analysis of connections between awareness score and use of preventive technologies. First we compared the means of the “Don’t knows”, “Non-users” and “Users” for each of the four preventive technologies, see Subsection 5.4. For all technologies, we saw a linear significant increase in mean awareness score from the “Don’t knows”, via the “Non-users” and to the “Users”. Anti-virus was the only technology not significant at the 1% level.

We saw an correlation between the awareness score and the four technologies at 0.081 for anti-spyware, 0.651 for anti-spyware, 0.436 for firewall and 0.599 for popup-blocker. The strong correlation coefficients for anti-spyware, firewall and popup-blocker indicate increase in use of the technologies as awareness rises. The exception is anti-virus, but this is again probably due to the fact that almost everyone use it, and thus the awareness score is not important.

From the regression analysis we found regression coefficients between awareness and anti-virus(0.032), anti-spyware(0.425), firewall(0.199), and popup-blocker(0.358). Again we see that use of anti-virus does not depend on and thus is not explained by awareness. For anti-spyware and popup-blocker the regression coefficients are rather high, close to the levels found by Dinev and Hu[7]. They found a regression coefficient of 0.47 between awareness and behaviour intention when it comes to adopting preventive technologies. This means that their awareness measurement explained 47% of the variance in behaviour intent.

Why awareness explains less of firewall-use than for anti-spyware and popup-blocker is

not clear after the analysis. Perhaps the fact that some of these technologies often are included in other software or hardware affect the regression coefficients. Another possible reason for the difference in explained use, could be that the threats anti-spyware and popup-blocker address are usually more visible for the user. Popups and other ads and similar as a result of e.g. adware are more likely to irritate the user, than a hacker or automated attack exploiting open ports and services on the PC, that the firewall is designed to protect against.

Difference between preventive technologies

From the comparison in trends among the different preventive technologies in subsection 5.4.4, we find that anti-virus have been adopted by pretty much everyone by now. The trends for our sample show that firewall is the technology first to be adopted, then popup-blocker and last anti-spyware.

Key knowledge for adoption

From the factor analysis, we determined trojans and spyware to be the most important threats to have knowledge on. The knowledge of popups, advertisements and toolbars as means of distributing malicious code was included in our awareness score and thus most important for the use of preventive technologies. And knowledge on anti-spyware, firewall and popup-blocker was the most important in the preventive technologies category. These variables gave us the best results in the regression analysis.

Threshold levels

From the Sections 5.4.3 and 5.4.4, we see that there are no clear threshold levels in use for our sample. There is a rather constant increase in use for both anti-spyware, firewall, and popup-blocker as awareness increases.

6.3 Interest in measures from Mypage

We see quite large interest in all services, see Section 5.5, between 50 and 75% are positive to the 7 different services asked about. But there is less interest in paying for these services, only 7,5% say they would be willing to pay for these services and another 25% say they might be willing to pay for the services. It should be noted that these results might be high due to the setting in which they were asked. These are positive questions asked after a lot of questions on threats and attacks from the Internet, and might have been biased by this. This was discussed in Section 3.3. So the actual use of any services if made available is likely to be lower than indicated in Section 5.5.

6.4 Methods

For the overall design of the thesis, we must say it performed mostly according to plan. We had some issues due to the time restrictions on the thesis, but did the best we could with what we had of resources available.

Instrument Unfortunately we did not have the time to validate the survey thoroughly before sending it out.The factor analysis is one example of this, where 17 of 26 questions where too unclear to be included in the further analysis. We are certain that we made the right decision in using only the valid parts of the survey for this analysis. Our work could then be looked at as more a foundation for further development of the measurement of awareness. Cronbachs alpha results was good

for both the initial awareness measure and after the reduction in the factor analysis. We had a alpha value of 0.908 after the factor analysis.

Recruitment In Section 6.1, we discussed the implications of our non-representative sample. But because of limits in resources, we did not have the opportunity to use a representative selection in this thesis. We got good response, approximately 11% of the estimated total number of people receiving the recruitment request.

Analyses The analyses of the data received from the survey has gone according to plan and we have come to valid conclusions from our data. We have used methods recommended from literature[15, 8, 29].

Tool With the number of responses we got, we had to move on to the more powerful statistical tool SPSS[48]. We utilized an evaluation version of the tool and it performed as hoped, after a short period of training.

7 Future work

Through our work with this thesis, we came across results and problems that would be interesting for future research.

Survey

Due to resource limits and the fact that our respondents were recruited on a voluntary basis, we felt that we could not make the survey too extensive. With the approximately completion time for the survey sat at 10 minutes, the information we could extract from the respondents was limited as well.

So the first thing we would recommend is to improve and expand the survey. A more extensive survey would provide additional details of knowledge and interest in the topics and better measure the awareness of the population. Perhaps a different type of survey is a better choice for measuring awareness, a quiz-form on the survey would definitively give a more accurate measure of knowledge. A combination of knowledge on and attitudes towards threats and protection might a good approach for the measurement tool. We have only asked people to rate their own knowledge and to actually test the knowledge would give more decisive results.

Further, the use of preventive section of the survey could be improved by actually scanning the user computer for malicious software infections and installed preventive technologies. This would be rather resource consuming, but one solution might be to cooperate with a security software supplier or similar.

Sample

What could also prove beneficial is to use a representative selection of the Norwegian population. Our selection differed from the population in age and education, and an representative selection would make the generalization to the population more accurate.

Analysis

In the factor analysis, we found a couple of other components including several of the variables not included in first component measuring knowledge on threats and preventive technologies. It would be interesting to look at variables 1, 2, 3, 4, 20, and 21 to investigate how these are related to the knowledge variables, as these variables measure habits and “practical” awareness(e.g. detecting if homepage changes).

The analysis of results could also be extended to include additional analysis techniques. Structural equation modeling (SEM) techniques such as linear structural relation(LISREL) and partial least square(PLS) are alternatives to the regression techniques we performed in this thesis.

Another area that would be interesting to more about is how previous use of one or more technologies affect the adoption of new technologies. It is likely that a positive experience with one preventive technology, e.g. that the number of irritating popups are reduced, would affect the adoption of other preventive technologies. But how great is

this effect and how does the order of preventive technology adoption affect the overall use of these technologies?

Other areas of further research

Further investigations of the open answers; Only three respondents mentioned privacy as a reason for using preventive technologies. It seems not many associate use of preventive technologies with privacy, which it would be interesting to find the reasons for.

8 Conclusions

A summary of the conclusions found by the work performed in this thesis. We return to the research questions in the introductory chapter, see Section 1.6.

Question 1: What is the awareness on the issues of privacy and threats among the Norwegian Internet users?

We did find a measurement for the awareness of our sample. But the awareness score found in this thesis is not very descriptive in itself, it needs to be seen in connection with the actual use. But our analysis found several key areas of knowledge that have an effect on the use of preventive technologies.

Question 2: To what extent is core preventive technologies utilized by Norwegian Internet users?

The investigations on use of preventive technologies showed us differences for the four technologies we were interested in. Anti-virus is almost standard as we see 94.1% use of it. We then found 72.3% use of a firewall and 66.5% use of a popup-blocker. But only 52.1% use anti-spyware software and since this has been pointed out as possibly the largest threat to privacy and information today[55], we argue that this is the most worrying result from the survey. Further work on charting use of preventive technologies should be conducted to investigate how much of the "Don't know"-group actually use anti-spyware, as this group includes almost 32% of the respondents.

Question 3: How does awareness affect the use of preventive technologies?

We have found similar connections between awareness and use of preventive technologies, as others[7]. But there are differences in how much awareness affects use of different technologies. Anti-virus is not affected as it used by virtually everyone, but awareness predicts well the use of both anti-spyware, firewall, and popup-blocker. It performs best in predicting use of anti-spyware(42.5%), then popup-blocker(35.8%) and last firewall(19.9%). The key areas of knowledge for use in our sample was trojans and spyware of threats, popups, advertisements and toolbars of the methods and anti-spyware, firewall and popup-blocker of the technologies.

Question 4: Is there a level of awareness that triggers the use of preventive technologies?

We compared the average use of the technologies at different levels of awareness, but found no clear threshold levels for any of the preventive technologies. In Figure 26 we see a steady increase in use as awareness increases in our sample, but we see a tendency of firewall being used first, then popup-blocker and last anti-spyware.

Question 5: How interested are potential users of web-based services like Mypage in educational material and security measures, if made available in affiliation with Mypage?

We asked about our respondents interest in making use of certain security services, if made available from Mypage. These services were guides to safe behaviour online, updated threat information, guides to installing, configuring and using preventive mea-

sures, vulnerability-check of PC, free preventive technology downloads, and online scan and removal of malicious code. We saw quite large interest in all services, between 50 and 75% are positive to the 7 different services asked about. But there is less interest in paying for these services, only 7,5% say they would be willing to pay for these services and another 25% say they might be willing to pay for the services.

Bibliography

[1] AOL/National Cyber Security Alliance. Online safety study 2004. 2004. http://www.staysafeonline.org/pdf/safety_study_v04.pdf, Last visited 28th May 2007.

[2] AOL/National Cyber Security Alliance. Online safety study 2005. 2005. http://www.staysafeonline.org/pdf/safety_study_2005.pdf, Last visited 28th May 2007.

[3] Neveen Farag Awad and Kristina Fitzgerald. The deceptive behaviors that offend us most about spyware. *Commun. ACM*, 48(8):55–60, 2005.

[4] Jon Bing. Personvern i faresonen. J.W. Cappelens Forlag AS, 1991. http://www.personvern.uio.no/pvpn/artikler/personvern_i_faresonen.pdf, Last visited 28th May 2007.

[5] Ragnar D. Blekeli and Knut S. Selmer, editors. *Data og personvern*. Universitetsforlaget, 1977.

[6] John W. Creswell. *Research design*. SAGE Publications, 2003.

[7] Tamara Dinev and Qing Hu. The centrality of awareness in the formation of user behavioral intention toward preventive technologies in the context of voluntary use. In *The Fourth Annual Workshop on HCI Research in MIS, Internation Conference of Information Systems (ICIS)*. http://sigs.aisnet.org/SIGHCI/Research/ICIS2005/SIGHCI_2005_Proceedings_paper_5.pdf, Last visited 28th May 2007.

[8] Terje Andreas Eikemo and Tommy Høyvarde Clausen, editors. *Kvantitativ analyse med SPSS*. Tapir Akademisk Forlag, 2007.

[9] EOS Gallup Europe. Data protection in the eu. 2003. http://ec.europa.eu/public_opinion/flash/fl147_data_protect.pdf, Last visited 28th May 2007.

[10] European Opinion Research Group. Data protection. 2003. http://ec.europa.eu/public_opinion/archives/ebs/ebs_196_en.pdf, Last visited 28th May 2007.

[11] Finjan Inc. Web security trends report - q1/2007. *http://www.finjan.com/Content.aspx?id=827#SecurityTrendsReport, Last visited 28th May 2007*, 2007.

[12] Arlene Fink. *The Survey Kit, Volume 2: How to ask survey questions*. SAGE Publications Inc., 1995.

[13] Free Software Foundation Inc. Pspp. http://www.gnu.org/software/pspp/, Last visited 28th May 2007.

[14] Lee A. Freeman and Andrew Urbaczewski. Why do people hate spyware? *Commun. ACM*, 48(8):50–53, 2005.

[15] Darren George and Paul Mallery. *SPSS for Windows Step by step, 6th edition.* Pearson Education, Inc., 2006.

[16] Beth Givens. Public records on the internet: the privacy dilemma. In *CFP '02: Proceedings of the 12th annual conference on Computers, freedom and privacy*, pages 1–7, New York, NY, USA, 2002. ACM Press.

[17] Randi Gjerde. Browser eavesdropping: how can we prevent our browsers from revealing out private information. Master's thesis, Gjøvik University College, 2005.

[18] Erik Griffin. Min side behovsstudie. Vindfang AS, 2005.

[19] Gustav Haraldsen. *Spørreskjemametodikk etter kokebokmetoden*. Ad Notam Gyldendal, 1999.

[20] Frank Helten and Bernd Fischer. What do people think about cctv? findings from a berlin survey. 2004. http://www.urbaneye.net/results/ue_wp13.pdf, Last visited 28th May 2007.

[21] Qing Hu and Tamara Dinev. Is spyware an internet nuisance or public menace? *Commun. ACM*, 48(8):61–66, 2005.

[22] IBM Internet Security Systems. X-force 2006 trend statistics, January 2007. http://www.iss.net/documents/whitepapers/X_Force_Exec_Brief.pdf, Last visited 28th May 2007.

[23] IBM Internet Security Systems. X-force threat insight quarterly, January 2007. http://documents.iss.net/ThreatIQ/ISS_XFTIQ_Q406.pdf, Last visited 28th May 2007.

[24] Justisdepartementet. Den almindelige norske straffelov(straffeloven), lov av 22 may 1902 no. 10. 1902-05-22, sist endret 2006-02-16. http://www.lovdata.no/all/hl-19020522-010.html, Last visited 28th May 2007.

[25] Justisdepartementet. Personregisterloven. 1978-06-09. http://www.lovdata.no/oll/nl-19780609-048.html, Last visited 28th May 2007.

[26] Justisdepartementet. Lov om behandling av personopplysninger (personopplysningsloven). 2000-04-14. http://www.lovdata.no/all/nl-20000414-031.html, Last visited 28th May 2007.

[27] Gregg Keizer. Dutch police crush big botnet, arrest trio. *http://www.techweb.com/wire/security/171204478, Last visited 28th May 2007*, 2005.

[28] Younghwa Lee and Kenneth A. Kozar. Investigating factors affecting the adoption of anti-spyware systems. *Commun. ACM*, 48(8):72–77, 2005.

[29] Paul D. Leedy and Jeanne Ellis Ormrod. *Practical research - Planning and design. 8th edition.* Pearson Education Inc., 2005.

[30] Ministry of Justice. Personal data filing system act, unofficial english translation of [25]. 1978-06-09. http://www.ub.uio.no/ujur/ulovdata/lov-19780609-048-eng.pdf, Last visited 28th May 2007.

[31] Ministry of Justice. Personal data act, unofficial english translation of [26]. 2000-04-14. http://www.ub.uio.no/ujur/ulovdata/lov-20000414-031-eng.pdf, Last visited 28th May 2007.

[32] Ministry of Justice and the Police. The general civil penal code, unofficial english translation of [24]. 1902-05-22, last changed 2005-12-21. http://www.ub.uio.no/ujur/ulovdata/lov-19020522-010-eng.pdf, Last visited 28th May 2007.

[33] Kevin D. Mitnick and William L. Simon. *The art of deception - Controlling the human element of security*. Wiley Publishing, Inc., 2002.

[34] Tove L. Mordal. *Som man spør får man svar*. TANO A.S., 1989.

[35] Norwegian Board of Technology. Holdninger til personvern (attitudes towards privacy). 2004. http://www.teknologiradet.no/Rapport_fokusgrupper_9-5lz.pdf.file, Last visited 28th May 2007.

[36] Tor Nygaard. *Skjemavett*. Kommuneforlaget, 2002.

[37] Arne Skauge og Personregisterlovsutvalget. Nou1997:19. et bedre personvern - forslag til lov om behandling av personopplysninger. 1997. http://www.regjeringen.no/Rpub/NOU/19971997/019/PDFA/NOU199719970019000DDDPDFA.pdf, Last visited 28th May 2007.

[38] Gunter Ollman. The phishing guide: Understanding and preventing phishing attacks. Next Generation Security Software Ltd, 2004. http://www.ngssoftware.com/papers/NISR-WP-Phishing.pdf, Last visited 28th May 2007.

[39] Gunter Ollman. The pharming guide: Understanding and preventing dns-related attacks by phishers. Next Generation Security Software Ltd, 2005. http://www.ngssoftware.com/research/papers/ThePharmingGuide.pdf, Last visited 28th May 2007.

[40] Robin Poston, Thomas F. Stafford, and Amy Hennington. Spyware: a view from the (online) street. *Commun. ACM*, 48(8):96–99, 2005.

[41] Privacy Rights Clearinghouse. A chronology of data breaches, Updated 2007. http://www.privacyrights.org/ar/ChronDataBreaches.htm, Last visited 28th May 2007.

[42] Inger-Anne Ravlum. Tøi-rapport 800/2005, processing of personal data in norwegian organisations. *http://www.toi.no/getfile.php/Publikasjoner/T%D8I%20rapporter/2005/800-2005/T%D8I-rapport-800-2005.pdf, Last visited 28th May 2007*, 2005.

[43] Inger-Anne Ravlum. Tøi-report 789/2005, pinning our faith on big brother ... together with all the little brothers? 2005. http://www.toi.no/getfile.php/Publikasjoner/T789-2005.pdf, Visited 28th May 2007.

[44] Paul Roberts. ISP telenor cripples zombie pc network. *http://www.infoworld.com/article/04/09/10/HNzombienetwork_1.html, Last visited 28th May 2007*, 2004.

[45] Ann Rudinow Saetnan, Johanne Yttri Dahl, and Heidi Mork Lomell. Views from under surveillance. public opinion in a closely watched area in oslo. 2004. http://www.urbaneye.net/results/ue_wp12.pdf, Last visited 28th May 2007.

[46] Dag Wiese Schartum and Lee A. Bygrave. *Personvern i informasjonssamfunnet*. Fagbokforlaget Vigmostad & Bjørke AS, 2004.

[47] Mark B. Schmidt and Kirk P. Arnett. Spyware: a little knowledge is a wonderful thing. *Commun. ACM*, 48(8):67–70, 2005.

[48] SPSS Inc. Spss for windows. http://www.spss.com/spss/, Last visited 28th May 2007.

[49] Thomas F. Stafford. Introduction. *Commun. ACM*, 48(8):34–36, 2005.

[50] Statistisk Sentralbyrå. Norsk mediebarometer: Andel som bruker internett en gjennomsnittsdag, etter kjønn, alder og utdanning. 2006. prosent., 2007. http://www.ssb.no/emner/07/02/30/medie/sa86/internett.pdf, Last visited 25th May 2007.

[51] Invar Tjøstheim, Kristin S. Fuglerud, Knut Boge, Ragni R. Arnesen, and Mette Langaas. Online-consumers and privacy. 2001. http://www.nr.no/~ingvar/Privacy-report979-2001.pdf, Last visited 28th May 2007.

[52] Survey Analytics LLC. Questionpro survey software. http://www.questionpro.com, Last visited 28th May 2007.

[53] Technical University of Berlin. Urban eye - on the threshold of urban panopticon?, 2001-2004. http://www.urbaneye.net, Last visited 28th May 2007.

[54] The faculty of Law Library, Unversity of Oslo. A collection of translated norwegian legislation. http://www.ub.uio.no/ujur/ulov/english.html, Last visited 28th May 2007.

[55] Roger Thompson. Why spyware poses multiple threats to security. *Commun. ACM*, 48(8):41–43, 2005.

[56] Samuel Warren and Louis Brandeis. The right to privacy. 1890.

[57] Webroot Software Inc. State of spyware q2 2006. *http://h30307.www3.hp.com/pdf/SOS_Q206_USA.pdf, Last visited 28th May 2007*, 2006.

[58] Alan F. Westin. *Privacy and freedom*. Atheneum, New York, 1967.

[59] Oleg Zaytsev. *Rootkits, Spyware/Adware, Keyloggers and backdoors: Detection and neutralization*. A-List Publishing, 2006.

[60] Xiaoni Zhang. What do consumers really know about spyware? *Commun. ACM*, 48(8):44–48, 2005.

A Articles from media

Here we give a few examples of articles from media, which proves that the threats discussed in this thesis are real and are becoming a serious problem. Both Skandiabanken, Nordea, Sparebank1 and DnBNOR banks has had unauthorized transactions on behalf of customers that has had their computers infected with trojans.

Examples from media:
http://www.dagensit.no/bedrifts-it/article992788.ece
http://www.dagensit.no/min-it/article1041609.ece
http://www.dagensit.no/finans/article1001853.ece
http://www.dn.no/forsiden/politikkSamfunn/article963881.ece
http://www.dagensit.no/min-it/article1041604.ece
http://www.dagensit.no/bedrifts-it/article1002109.ece

At least Skandiabanken and DnBNOR has changed their security measures as a result of these attacks.

Links to articles from the media and similar, concerning privacy breaches at governmental bodies and other public institutions:

Mandal municipality:
http://www.nrk.no/nyheter/distrikt/ sorlandet/1.1728754
Ålesund municipality:
http://www.smp.no/default.asp?page=1003&item=865366,1&lang=1
Vest-Agder municipality:
http://www.nrk.no/nyheter/distrikt/sorlandet/1.1728754
Elverum municipality:
http://www.ostlendingen.no/apps/pbcs.dll/article?AID=/20070228/NYHETER/70228005
University of Oslo:
http://foreninger.uio.no/akademikerne/brevet.html
BNbank:
http://www.bt.no/innenriks/article300581.ece

All link destinations in Norwegian and all links last visited 28th June 2007.

B Survey questions

In the following pages we included the survey questions approximately as the appeared for the respondents. The Norwegian text is the exact same, but in the conversion from the web form to pdf we chose to compress it and put several questions on each page.

We also numbered the questions consecutively as we needed to refer to a few of them in chapter 5.

Hei!

Tusen takk for at du tar deg tid til å hjelpe meg med masteroppgaven!
Deltagelsen er helt frivillig og dersom du er ukomfortabel med å svare på noen av spørsmålene, kan du når som helst avslutte. Det vil ta ca. 10 minutter å fullføre undersøkelsen. Alle svar behandles konfidensielt og resultater fra undersøkelsen vil kun bli presentert på et overordnet nivå.

Med vennlig hilsen,Freddy Lønne Andreassen

Om deg selv

Question / Spørsmål 1:
Kjønn:
- Mann
- Kvinne

Question / Spørsmål 2:
Alder:
- Under 18
- 18 - 24
- 25 - 34
- 35 - 44
- 45 - 54
- Over 55

Question / Spørsmål 3:
Hvilket postnummer har hjemstedsadressen din?

Question / Spørsmål 4:
Utdannelse:
- Ungdomsskole
- Videregående
- Høyskole/universitet

Question / Spørsmål 5:
Arbeidssituasjon:
- Student
- Privat ansatt
- Offentlig ansatt
- Selvstendig næringsdrivende
- Pensjonist
- Arbeidsledig
- Annet, inkl. hjemmeværende, førstegangstjeneste, trygdet

Question / Spørsmål 6:
Dersom du svarte privat eller offentlig ansatt: Hvilken yrkestype har du?
- Admin./konomi, kontor og jus
- Handel kundeservice, restaurant og reiseliv
- Industri, bygg/anlegg, håndverk og verkstedsarbeid
- Transport, Logistikk, kommunikasjon og IT
- Jord-/skogbruk, fiske og matproduksjon
- Kultur, religiøst arbeid, idrett og informasjonsformidling
- Skole, fritid, undervisning og forskning
- Helse, omsorg, medisin og biologi
- Service- og sikkerhetsarbeid
- Andre

Question / Spørsmål 7:
Hvor mange års erfaring har du med bruk av PC?

- Under 2
- 2 - 5
- 6 - 10
- 11 - 15
- 16 - 20
- Over 20

Question / Spørsmål 8:
Hvor mange års erfaring har du med bruk av Internett?

- Under 2
- 2 - 5
- 6 - 10
- Over 10

Surfevaner

Question / Spørsmål 9:
Hvorvidt klikker du på reklamebannere eller reklamer i sprett-opp-vinduer (popups)?

- Ofte
- Nå og da
- Sjelden
- Aldri

Question / Spørsmål 10:
Installerer du programmer du blir anbefalt via reklamebannere og sprett-opp vinduer(popups)?

- Ofte
- Nå og da
- Sjelden
- Aldri

Spørsmål 11:
Når man vil installere programvare, må man som regel godta en sluttbrukeravtale før installasjonen starter. Leser du hele denne sluttbrukeravtalen før du installerer?

- Alltid
- Ofte
- Sjelden
- Aldri

Spørsmål 12:
Forstår du alt innhold i disse sluttbrukeravtalene?

- Ja
- Bare delvis
- Nei
- Leser ikke

Kunnskaper

Question / Spørsmål 13-18:
I hvilken grad har du kjennskap til følgende trusler/angrep mot en PC som er tilkoblet Internett og brukes til surfing?

	Ikke hørt om.	Kjenner ikke detaljer.	Kjenner til dette.	Har full oversikt og beskytter meg.
Virus	☐	☐	☐	☐
Trojaner	☐	☐	☐	☐
Spionprogram (Spyware)	☐	☐	☐	☐
(Keylogger)	☐	☐	☐	☐
(Phishing)	☐	☐	☐	☐
Sosial entrepenørkunst(Social engineering)	☐	☐	☐	☐

Question / Spørsmål 19-25:
Følgende metoder brukes av truslene/angrepene ovenfor, for å angripe en PC. I hvilken grad kjenner du til disse metodene?

	Ikke hørt om.	Kjenner ikke detaljer.	Kjenner til dette.	Har full oversikt og beskytter meg.
En link i en e-post	☐	☐	☐	☐
Vedlegg til e-post	☐	☐	☐	☐
Popup-vinduer	☐	☐	☐	☐
Reklame	☐	☐	☐	☐
Sluttbrukeravtaler	☐	☐	☐	☐
Verktøylinjer (Toolbars)	☐	☐	☐	☐
(Drive-by downloading)	☐	☐	☐	☐

Question / Spørsmål 26-27:
Vet du hva følgende sikkerhetsinformasjon fra nettleseren(f.eks. Internet Explorer, Firefox eller Opera) betyr?

	Ja	Nei	Vet ikke
Hengelåsen til høyre i adressefeltet i nettleseren	☐	☐	☐
Sertifikatet du kan se ved å trykke på hengelåsen	☐	☐	☐

Hendelser

Question / Spørsmål 28:
Hvor mange popup-vinduer får du anslagsvis i uka, mens du surfer på din PC?

- Under 5
- 5 - 10
- 11 - 15
- 16 - 20
- 21 - 25
- Over 25

Question / Spørsmål 29:
Har startsiden i nettleseren din forandret seg automatisk?

- Ja
- Nei
- Vet ikke

Question / Spørsmål 30:
Har du blitt omdirigert til en annen søkemotor automatisk, når du har gått inn på en søkemotor på Internett (som f.eks. Google eller Kvasir)?

- Ja
- Nei
- Vet ikke

Sikkerhetstiltak

Question / Spørsmål 31:
I hvilken grad mener du din PC er tilstrekkelig/utilstrekkelig sikret for det den brukes til?

- Tilstrekkelig sikret
- Delvis sikret
- Utilstrekkelig sikret
- Vet ikke

Question / Spørsmål 32-35:
I hvilken grad har du kjennskap til følgende sikkerhetstiltak for en PC:

	Ikke hørt om.	Hørt uttrykket.	Vet hvordan det brukes.	Kan installere og konfigurere.
Anti-virus	❑	❑	❑	❑
Anti-spyware	❑	❑	❑	❑
Brannmur (firewall)	❑	❑	❑	❑
Popup-blokkerer	❑	❑	❑	❑

Question / Spørsmål 33:
Bruker du anti-virus programvare på din PC?
Ja
Nei
Vet ikke

Question / Spørsmål 33a:
Hvorfor begynte du å bruke anti-virus programvare?

Question / Spørsmål 33b:
Hvorfor bruker du ikke anti-virus programvare?

Question / Spørsmål 34:
Bruker du anti-spyware programvare på din PC?
Ja
Nei
Vet ikke

Question / Spørsmål 34a:
Hvorfor begynte du å bruke anti-spyware programvare?

Question / Spørsmål 34b:
Hvorfor bruker du ikke anti-spyware programvare?

Question / Spørsmål 35:
Bruker du brannmur (firewall) på din PC?
Ja
Nei
Vet ikke

Question / Spørsmål 35a:
Hvorfor begynte du å bruke brannmur (firewall)?

Question / Spørsmål 35b:
Hvorfor bruker du ikke brannmur (firewall)?

Question / Spørsmål 36:
Bruker du popup-blokkerer på din PC?
Ja
Nei
Vet ikke

Question / Spørsmål 36a:
Hvorfor begynte du å bruke popup-blokkerer?

Question / Spørsmål 36a-a:
Har du merket en nedgang i antall sprett-opp-vinduer (popups) etter du begynte å bruke en blokkerer?

- Ja
- Nei
- Vet ikke

Question / Spørsmål 36b:
Hvorfor bruker du ikke popup-blokkerer?

Question / Spørsmål 37:
Hvor ofte oppdaterer du programvare på din PC (hovedsaklig sikkerhetstiltakene nevnt i tidligere spørsmål, i tillegg til operativsystem)?

- Hver gang nye oppdateringer er tilgjengelige
- Jevnlig
- Sjelden
- Aldri
- Vet ikke

Minside på Norge.no

Question / Spørsmål 38-43:
Hvorvidt vil du benytte deg av følgende tjenester, dersom de blir gjort tilgjengelige i forbindelse med Minside?

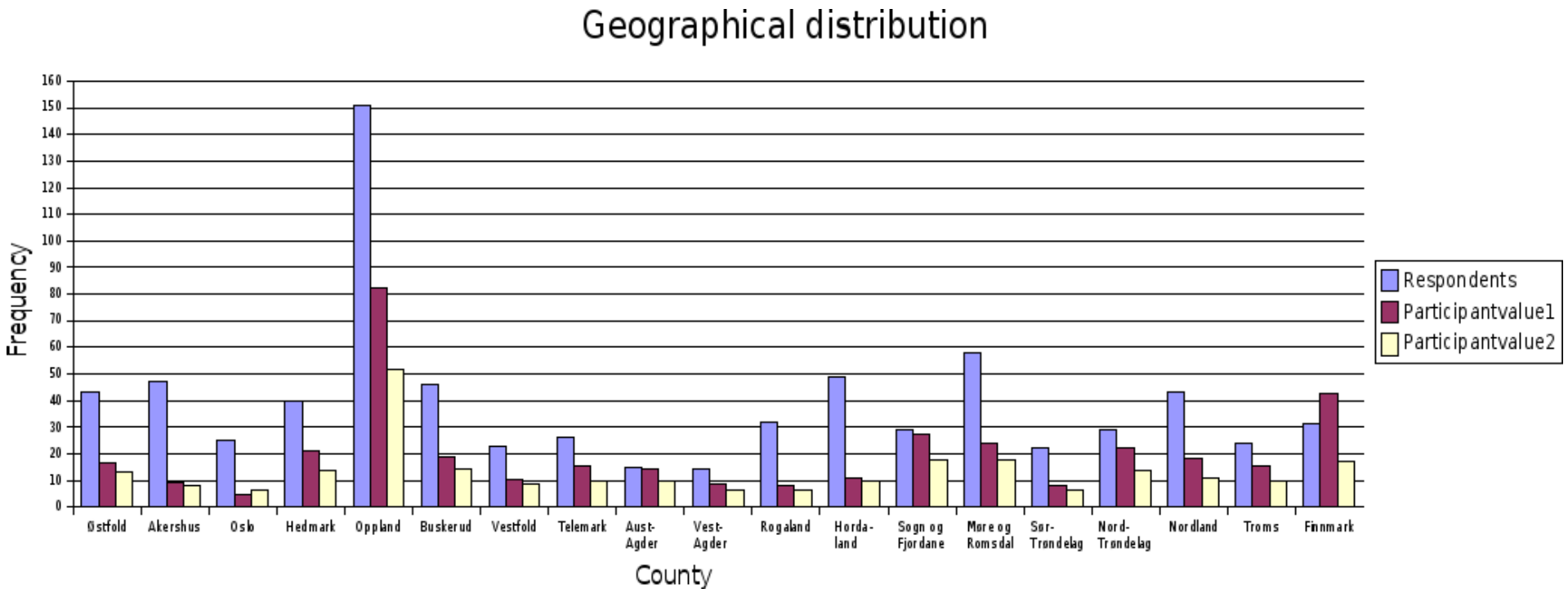
	Ja.	Mest sannsynlig.	Sannsynligvis ikke.	Nei.	Vet ikke
Veiledning i sikker surfing.	❑	❑	❑	❑	❑
Oppdatert informasjon om de vanligste trusler/angrepmot en PC som er tilkoblet Internett.	❑	❑	❑	❑	❑
Veiledning i installasjon, konfigurasjon og brukav sikkerhetstiltak.	❑	❑	❑	❑	❑
Sårbarhetsanalyse av din PC.	❑	❑	❑	❑	❑
Gratis sikkerhetstiltak i form av nedlastbarprogramvare.	❑	❑	❑	❑	❑
Nettbasert sjekk og fjerning av virus ogspionprogrammer på din PC.	❑	❑	❑	❑	❑

Question / Spørsmål 44:
Ville du kunne tenke deg å betale en avgift for disse tjenestene, enten som en engangsavgift eller som en abonnementsløsning?

- Ja.
- Mest sannsynlig.
- Sannsynligvis ikke.
- Nei.
- Vet ikke.

C Geographical distribution

This is the geographical distribution of our sample, see next page for illustration. The frequency scale on the left is only applicable to the “Respondents” values. The “Participantvalue1” is determined by comparing number of respondents from one county with the total number of respondents from that county. The “Participantvalue2” is determined by comparing the number of respondents from one county with the number of employees in the public administration for that county. The participant values are only meant to be comparable between the different counties and not in any other way. The “Participantvalue2” is of course biased by the number of students form GUC answering the survey.


Geographical distribution
Frequency
160
150
140
130
120
110
100
90
80
70
60
50
40
30
20
10
0
Østfold
Akershus
Oslo
Hedmark
Oppland
Buskerud
Vestfold
Telemark
Aust-Agder
Vest-Agder
Rogaland
Horda-land
Sogn og Fjordane
Møre og Romsdal
Sør-Trøndelag
Nord-Trøndelag
Nordland
Troms
Finnmark
County
Respondents
Participantvalue1
Participantvalue2

D Recoding of data

This section includes the complete recodes performed before the statistical analysis, see Table 6 for details.

Question number	Answer alternative	Old value	New value
11	“Ofte”(Often)	1	4
	“Nå og da”(Sometimes)	2	3
	“Sjelden”(Rarely)	3	2
	“Aldri”(Never)	4	1
12	“Ofte”(Often)	1	4
	“Nå og da”(Sometimes)	2	3
	“Sjelden”(Rarely)	3	2
	“Aldri”(Never)	4	1
26	“Ja”(Yes)	1	3
	“Nei”(No)	2	2
	“Vet ikke”(Don’t know)	3	1
27	“Ja”(Yes)	1	3
	“Nei”(No)	2	2
	“Vet ikke”(Don’t know)	3	1
29	“Ja”(Yes)	1	2
	“Nei”(No)	2	2
	“Vet ikke”(Don’t know)	3	1
30	“Ja”(Yes)	1	2
	“Nei”(No)	2	2
	“Vet ikke”(Don’t know)	3	1

Table 6: Recoding of answer alternatives
English translation in parentheses

E Factor analysis

Component Matrix[a]

	Component						
	1	2	3	4	5	6	7
ClickAds	,156	,036	-,314	,564	,394	-,095	,205
InstallPrograms	,148	,042	-,284	,622	,273	-,136	,310
ReadEULA	-,082	,070	,823	,281	,177	,058	-,041
UnderstandEULA	,121	,082	,811	,232	,223	,037	-,074
KnowVirus	,635	,380	,037	,095	-,081	,000	-,087
KnowTrojan	,734	,127	-,025	-,101	,169	-,273	-,160
KnowSpyware	,784	,139	,003	-,084	,134	-,260	-,115
KnowKeylogger	,768	-,410	,021	,026	-,023	-,168	,004
KnowPhishing	,756	-,360	-,028	,028	,045	-,166	,003
KnowSocialengineering	,688	-,486	,052	,079	-,101	-,174	,051
KnowMethodLink	,685	,356	-,029	,196	-,279	,276	-,018
KnowMethodAttach	,709	,399	-,080	,187	-,247	,238	-,060
KnowMethodPopup	,760	,233	-,054	,103	-,243	,213	,044
KnowMethodAd	,735	,061	-,052	,150	-,265	,159	,112
KnowMethodEULA	,625	-,302	,217	,097	-,338	-,016	,101
KnowMethodToolbar	,764	-,276	,002	,037	-,178	,014	,081
KnowMethodDriveby	,609	-,407	,176	,031	-,265	-,074	,108
KnowPadlock	,593	-,306	-,142	-,013	,372	,459	-,286
KnowSertificate	,614	-,347	-,114	-,009	,364	,428	-,268
ChangedHomepage	,333	,056	,078	-,285	,165	,114	,511
RedirectSearchengine	,255	-,013	,127	-,339	,224	,312	,586
KnowAnti-virus	,687	,335	-,007	-,127	,147	-,129	-,081
KnowAnti-spyware	,797	,121	-,017	-,187	,160	-,211	-,116
KnowFirewall	,755	,211	,003	-,095	,086	-,071	-,028
KnowPopupblock	,759	,149	-,017	-,251	,136	-,030	,001
Updates	,532	,263	,107	-,133	,118	-,186	,154

Extraction Method: Principal Component Analysis.

a. 7 components extracted.

Figure 34: Factor component matrix

Comments

Factor 1: We note that variables ClickAds, InstallPrograms, ReadEULA, UnderstandEULA, ChangedHomepage and RedirectSearchengine give low factor loads. All other variables have high loads on this factor, indicating that this is factor mainly measure knowledge.

Factor 2: Unclear factor, no factor loads above 0.5 on any variable.

Factor 3: Variables ReadEULA and UnderstandEULA loads very high on this factor, no other variable give a good factor load.

Factor 4: Variables ClickAds and InstallPrograms has good loads on this factor, none of the other factors have a good load on this factor.

Factor 5: Unclear factor, no factor loads above 0.5 on any variable.

Factor 6: Unclear factor, no factor loads above 0.5 on any variable.

Factor 7: Variables ChangedHomepage and RedirectSearchengine has factor loads above 0.5 on this factor and no other variable has a factor load close to this.