Towards Privacy Managment of Information Systems

Vidar Drageide
University of Bergen
Department of Informathics
June 2, 2009

NoWires Research Group
www.nowires.org

Preface

This masters thesis provides insight into the concept of privacy. It argues why privacy is important, and why developers and system owners should keep privacy in mind when developing and maintaining systems containing personal information. Following this, a strategy for evaluating the overall level of privacy in a system is defined. The strategy is then applied to parts of the cellphone system in an attempt to evaluate the privacy of traffic and location data in this system.

The thesis was written at the University of Bergen, Department of Informatics, and was supervised by Professor Kjell Jørgen Hole at the Selmer Center.

Acknowledgements

First and foremost I would like to thank my supervisor Professor Kjell Jørgen Hole, not only for his excellent guidance through the work with my thesis, but also for the opportunity to spend the last two years in the NoWires Research Group. During these years I have had the pleasure of working alongside brilliant people, who not once were too busy to answer my questions, gave me feedback, and included me in their discussions. It has been a privilege to make your acquaintance.

My work at the university would not have been possible had it not been for my common law spouse Kjersti, her love, support, and not least, patience; extended through these years, and especially the last couple of weeks. I should also say thank you to my son Brynjar, for sleeping quietly during the nights and his special talent for cheering me up after long days of work. I would not have made it without your help.

Thanks is also extended to all my fellow students who filled these five years at the university with joyful memories. At last I would like to thank my family for their support and love.

Contents

List of Tables

List of Figures

Chapter 1

Introduction

"You have zero privacy anyway, get over it."
Scott McNealy, CEO Sun Microsystems, 1999

In 1969, an American research project connected University of California, Los Angeles and Stanford Research International together and formed the start of the Advanced Research Projects Agency Network (ARPANET). This marked the beginning of the Internet as we know it today, and the academic network quickly grew in size. In 1991, when the European Organization for Nuclear Research (CERN) went public with their WorldWideWeb project the Internet started to change from something academics and the military were bothered with, to a natural part of our everyday lives. According to the Internet Systems Consortium the number of hosts have grown from just above one million in 1993, to above staggering 600 000 000 [1].

With the growth of the Internet, more and more services have become available online. Today people use online services for various tasks, anything from handing in homework, keeping in touch with friends, online shopping, banking, dating, the list just goes on. Additionally, new services keep emerging, for quite a few years Norwegians have been able to deliver their tax returns online, the Norwegian government even has a project on electronic voting, where one of the goals is to facilitate remote electronic voting [2, 3]. One online technology that has gained a lot of steam lately is social networking. Services like Facebook and Myspace let people build a virtual copy of their social network. The users are encouraged to connect to people they know from the real world, disclose how they know them, and share different types of information.

But as more and more services are brought online, the amounts of information about individuals residing on the Internet increases. If someone was

to fetch information from the different online services, they would probably be able to tell quite a lot about the individuals registered.

This thesis considers the privacy of such information residing in large information systems. Alongside the great technological development we have seen during the last decades, lawmakers have tried to keep up by introducing new legislation in an effort to regulate how such information should be protected. But legislation alone does not lead to compliance, and when developing information systems, that perhaps are to be used over long periods of time, it is hard to ensure that the systems keep personal information safe; much like building secure systems is hard.

Throughout this thesis the reader will be given an introduction to privacy, and to why privacy is important. After this introduction, the thesis attempts to define an approach that may be employed in order to evaluate the overall privacy level of large information systems. The idea is that such a structured approach will be beneficial both to companies and their consumers. Companies are able to verify that they are operating in accordance to legal requirements, and identify points in their systems where privacy is suffering, while consumers are ensured that systems are designed with their privacy in mind.

In addition to the suggested approach, the thesis also contains a privacy evaluation of personal information residing in the mobile telephone network. The evaluation is included as an example of how the suggested method may be used to identify and highlight areas where privacy is not properly taken care of. As a part of this review, a program for mobile phones was written to exemplify how geolocalisation of mobile devices can be done.

Privacy is, as we shall se in Chapter 2, a difficult concept to define, and an area where the amount of subjectivity is large. A certain degree of privacy may be good enough for one person, while appalling to another. As a result, I have chosen to write parts of this thesis using first person singular. This is a conscious choice, that is supposed to stress the subjectivity of the elements discussed in the sections where this is the case, and that some of the statements made are hypotheses made by the author, and not necessarily universal truths.

The intended audience of this thesis is anyone that has an interest in privacy, but it is specially meant for computer scientists who are developing systems processing personal information. The hope is that the method suggested can be a valuable tool in order to ensure the privacy of those who will one day use the system. I also think that the method described can be useful for organizations who have systems that are currently containing personal information, and that applying the method to a live system will make it possible to identify potential problems; or at the very least give the

organization a reassurance that they are doing a good job maintaining the privacy of their customers.

1.1 Structure of the Thesis

The following paragraphs gives an outline of the chapters in the thesis.

Chapter 2

Chapter 2 serves as an introduction to privacy. The chapter starts by trying to define privacy and continues to argue why privacy is needed, and why it is important to factor in privacy while developing systems that contain information about people. It then tries to explain why the erosion of privacy continues. Many important terms are also defined throughout this chapter, and it serves as a basis for the rest of the thesis.

Chapter 3

In Chapter 3, the reader is given an introduction to risk management, followed by a discussion about metrics. The chapter continues to develop a method for analyzing the overall privacy in an information system, using risk management as a source of inspiration.

Chapter 4

In this Chapter, I have applied my own process to parts of the mobile phone network, and tried to analyze it with respect to privacy of personal information contained within the system. The Chapter also contains an introduction to the most important elements in a mobile phone network, as an understanding of it is needed to say something about the level of privacy it provides.

Chapter 5

Chapter 5 contains a summary of the thesis and some conclusions. It also contains a small section with my reflections and suggestions for future work on this topic.

Appendix A

In Appendix A a sample midlet I wrote to determine the geographic position of mobile users is included. The midlet is Sony Ericsson specific, but the

strategy used for determining the position of a phone is applicable to all phones where access to network information is granted through API-calls.

Appendix B

Appendix B contains a sample of the information that was given to me by the phone company I am currently a customer of. It is included to illustrate some of the data your service providers has access to.

Chapter 2

Privacy

"Everyone has the right to freedom of opinion and expression; this right includes freedom to hold opinions without interference and to seek, receive, and impart information and ideas through any media and regardless of frontiers."
Universal Declaration of Human Rights, §19

In this chapter I introduce the concept of privacy and define some basic terms that enable us to dig deeper into the area from a computer science point of view. Then I argue why we need privacy, despite the fact that it might seem tempting to trade it in for other short term benefits. Next, I explain why it is necessary to measure the level of privacy in systems in order to evaluate it. The Chapter ends with an outline of one approach used by the American government to ensure that privacy is preserved in information systems.

2.1 What is Privacy

Although privacy is something everyone seems to have an opinion on, and many have strong feelings about, giving a crisp definition of privacy is not a simple task. It is a task that has puzzled philosophers and scholars for centuries without yet reaching a consensus. The modern debate around privacy surfaced in 1890 after the publication of a famous paper titled "The Right to Privacy" by the two american lawyers Samuel D. Warren and Louis D. Brandeis, where they argued about the existence of a right to privacy [4]. Back then, as today, privacy advocates were worried that emerging technologies would increase the overall exposure of people's private lives and undermine

their privacy. Warren and Brandeis argued that a fundamental right to privacy existed and that it should be granted protection by the juridical system.

Privacy has different meanings in different settings, and is more of an umbrella term for different personal rights. Usually, the term privacy is used in two different senses. Physical privacy, thought of as a right to prevent intrusion into ones physical space, such as a home, as well as the right to seclude oneself from others. The other type of privacy is informational privacy, seen as the right to seclude and control data about oneself.

The quote at the beginning of the Chapter is article 19 in the International Human Rights Charter, and it is interpreted to include the right to privacy as a human right. Even though considered to be a human right, very few nations include privacy as a part of their constitution, Norway is not an exception, although a recent study initiated by the Norwegian government recommended that privacy should be given its own constitutional protection [5, Ch. 19]. Today privacy in many countries is protected through a number of laws, regulations, and juridical precedence. Searching for the perfect definition of privacy and giving a thorough overview of how it is legally protected around the world is far beyond the scope of this thesis, for a good overview of privacy, and how privacy is protected by laws and agreements the reader is referred to the human rights organisation Privacy International and their overview of privacy [6].

2.2 Do We Really Need it?

As the quotation at the beginning of Chapter 1 wrote "You have zero privacy—Get over it." This statement from Scott McNeally, former CEO at Sun Microsystems, was given during a product launch in 1999. With discouraging statements like this, from a high-ranking official in one of the large multinational corporations within the computer industry, it might seem like privacy was something that only existed back in the 20th century. However, in the next sections I will argue that this is not the case, and at the very least there are a few systems where everyone can agree that privacy is needed.

In most discussions between privacy advocates and those who want to introduce some kind of privacy invasive technology, the argument "if you have nothing to hide then what are you afraid of?" is brought into play. The question implies that privacy is about hiding something that is wrong or illegal. The battle for privacy is rather a battle between the more fundamental issue of freedom versus control, and not a battle between privacy and security [7]. Consequently, privacy is not about hiding something wrong, but about having the freedom to seclude personal information or oneself if one

desires to.

Most types of crime could probably be prevented or discovered if we just introduce enough measures of controlling the actions of individuals and groups. Again it breaks down to which type of society we want to live in. The famous book "Nineteen Eighty Four" by George Orwell [8] portrays a society controlled by "the party." Everyone is under constant surveillance and any action deemed wrong, in one way or another, may lead to getting arrested by the "thought police." A more recent book of fiction is "Little Brother" by Cory Doctrow [9]. The book brings us into the life of a young boy that takes on the Department of Homeland Security, after they react to a terrorist attack by deploying all sorts of surveillance techniques, undermining the first amendment.

Of course these books are merely works of fiction, but for several years there have been initiatives for automatic profiling of airline passengers in the US through the earlier CAPPS program, and the suggested CAPPS II program [10]. As a side note, most of the techniques deployed by the Department of Homeland Security in the book by Doctrow, already exist, and are used in large scale on a day-to-day basis. The difference between fiction and the real world lies in how the technologies are used.

The British Home Office has recently released a report which states that automatic profiling is of limited use, partly due to a large number of false positives [11].

Following from the events of 9/11, but also from the continuous technological advances that would have appeared anyway, it is fairly safe to say that the amount of privacy a person enjoys in the western world today has not increased during the last decade. Numerous CCTV systems have been deployed, border-controls have been intensified, electronic passports and other ID-card schemes have been deployed, and these are just a few examples of large-scale systems that to some degree facilitate surveillance and undermine people's privacy. New emerging technologies may be used in different attempts at creating a safer society, although they might be very privacy invasive. For such technologies it can be argued that the individuals right to privacy outweighs the benefits that the specific technology is able to deliver, and that society is best of with them not deployed.

Many see privacy as a fundamental requirement for a well functioning democratic society. The term democracy is interpreted differently all around the world, but the foundation is a form of government where the people hold the power, and everyone is granted fundamental rights and freedoms. Privacy includes rights to gather information and to decide for oneself, and is thus necessary for something as basic as a free democratic election.

In a talk given at the German Chaos Computer Club Camp in 2007, the

philosopher Sandro Gaycken argues that in many of the discussions between privacy advocates and others arguing for more surveillance, privacy advocates are using arguments with soft values that fail to persuade people toward wanting more privacy [12]. Claims as "I don't like being watched" becomes insignificant in comparison with "This technology has shown to reduce serious and or violent crimes this much." Gaycken presents arguments advocating privacy and place these in three areas, psychological, sosioeconomical, and technopolitical consequences of surveillance. In the following paragraphs, I will present some of his arguments of why and how surveillance technologies can affect society in different ways.

2.2.1 The Good, the Bad, and Me?

In many situations surveillance equipment is used to enforce laws, and to prevent certain actions, ranging from preventing vandalism to enforcing speed limits on roads. Nevertheless, the way these laws are being enforced reflects a set of values that are believed to be important, at least by those spending money on these systems. A known "problem" in science is the observer effect, by observing something you affect it. In psychology, the effect is referred to as the Hawthorne effect. While studying the effectiveness of workers it was realized that productivity was higher when the workers knew they were under evaluation [13]. Gaycken argues that people who know they are under surveillance may be influenced by the surveillance in the same way. He further argues that people will somehow react to surveillance by trying to behave in accordance to, or in opposition to, what they think the observer wants.

A society where surveillance is conspicuously present may result in a monoculture of values. Individuality, the ability to judge ethically from own beliefs and opinions might be replaced with the capability to decide whether or not an action is in accordance with established values. Gaycken compares this tendency with studies done on children with overprotective mothers. Such studies have shown that these children get indecisive and very dependent of their mothers. They have poor ethical competence, and are behaving either like their mother wants them to, or in defiance with their mothers' wishes. Of course this is something that happens in an extreme setting, but just how much the current level of surveillance affects us as individuals is unknown, and do we really wish for a society where the amount of surveillance affects our daily behavior? I personally believe his arguments carry water, and that they should be the foundation of a discussion about privacy and surveillance.

Another problem with a surveillance infrastructure, and the resulting lack of privacy, that Gaycken brings up is how such an infrastructure facilitates

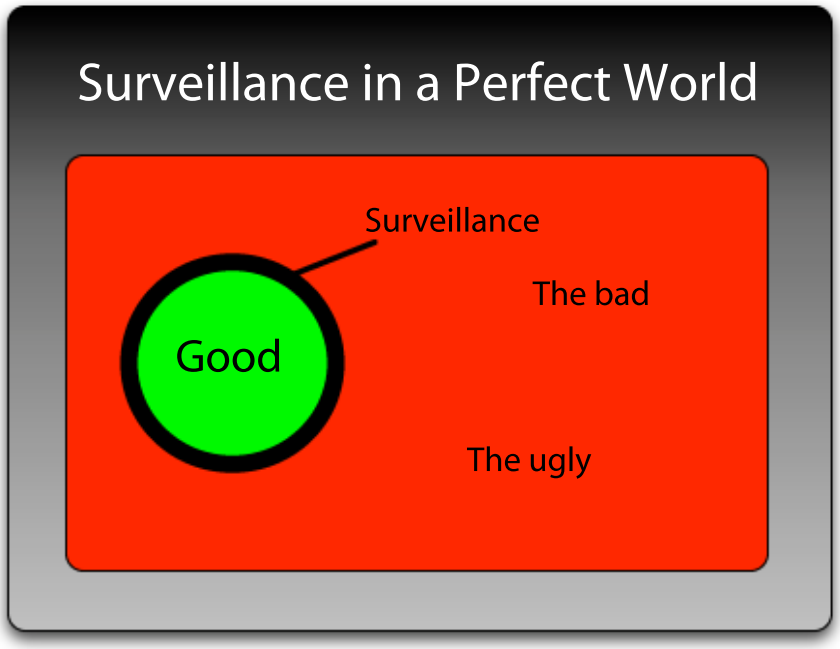


Figure 2.1: How surveillance is used to protect values in a perfect world.

the growth of a new class society based on automatic profiling.

People and organizations wanting to apply some form of surveillance, be it camera surveillance or large scale logging and processing of information, of course never say that they have bad intentions. They want to use technology to protect what they, and generally most of society, consider to be some kind of *good* against some kind of *evil*. This is illustrated in Figure 2.1 where surveillance is used to protect *good* from the *bad* and the *ugly*.

But who decides what is good and what is not? Generally people agree that most forms of crime are bad and should thus be fought; although most of us also include the notion of proportionality, meaning that society should not use any means possible in order to fight any kind of crime. Methods used to fight crime has to be proportional to how bad we consider the specific act to be.

A large corporation however does not necessarily care about all crime, what is important to them is generally to maximize their profit, and to avoid engaging in business with people and organizations that are not profitable to them. After all that is what the stockholders are expecting. The shift in priorities turns the table a bit in terms of surveillance. If such an organization is the owner of a surveillance infrastructure, suddenly *good* in the sense society sees it, has been substituted by *profit*, and *evil* has been substituted with

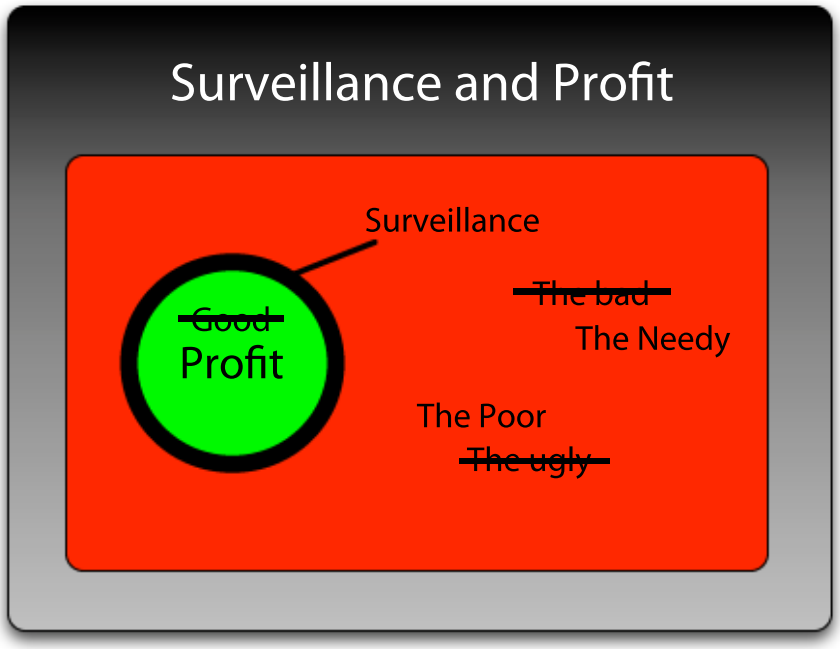


Figure 2.2: What if the owner seeks to maximize their profits?

the *poor and the needy*. Thus turning the use of surveillance away from the ideal situation illustrated by Figure 2.1, into a less ideal situation depicted in Figure 2.2; illustrating how the incentives of system owners influence how a certain technology is used.

As more surveillance technologies, and other privacy pervasive systems, are introduced in the name of "war against terror" or "for the sake of the children" in the western world, it gets harder to argue against the deployment of surveillance and privacy invasive systems in countries whose governments more resembles dictatorships. How can we argue that it is okay for us to deploy such systems, but not for them? In countries ruled by dictatorships, or any other totalitarian form of state, the use of privacy invasive systems are further shifted from the original intentions. Here systems may be employed to protect the elite, or the dictator, against everyone else as illustrated in Figure 2.3.

Some of the examples Gaycken uses involves taking the argumentation to the extreme, and exemplify worst case scenarios, but nonetheless he has a lot of good points and for those interested in privacy the German Chaos Computer Club has a video of his talk available at [12]. Even if his scenarios are extreme, I have not been able to find any research talking about how the current level of surveillance is affecting us.

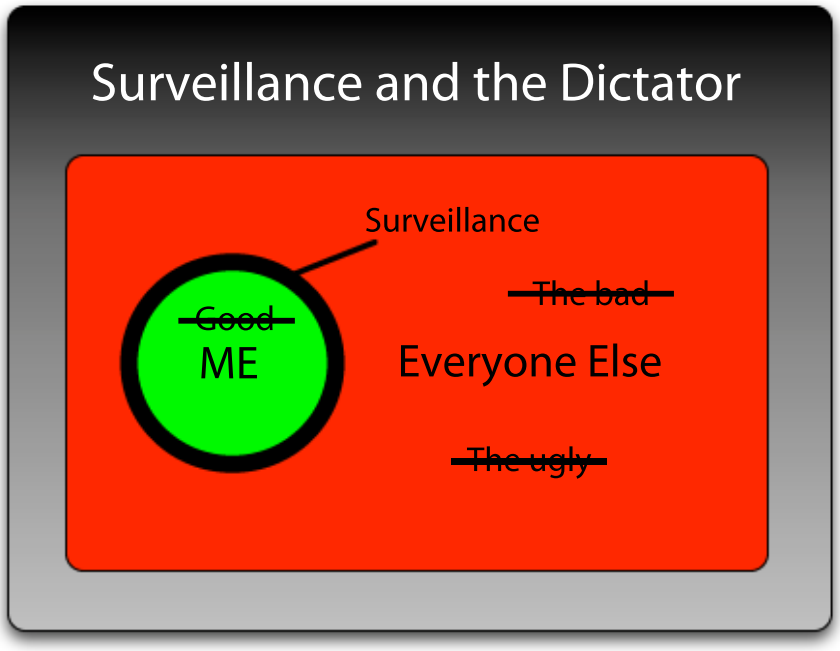


Figure 2.3: What if the system is controlled by a dictator?

2.3 Why Does the Privacy Erosion Continue?

As we saw in the previous Section, there exists solid and well-fundamented reasons to preserve privacy, and some of them seem very important. In particular, privacy is a necessity for a functioning democracy, and its absence can affect our ability to judge ethically. We also saw that the widespread deployment of privacy invasive systems seems to continue at full speed. In this section, I will try to give some insights as to why such systems keep on emerging.

The English word "surveillance" came from French and literally means to watch over or care for. Georg Apenes, the director of the Norwegian Data Inspectorate once said that big brother walks hand in hand with big mother [14]; implying that some of the systems undermining privacy are there to control those under surveillance, while other systems exist to protect people from known and unknown dangers they may expose themselves to, as a caring mother would do. He argued that big mother represents the democratic state, and its overeagerness to care for its citizens, and making sure that they adhere to what the state considers their best interests. So while some systems are there as necessary safeguards to prevent people from cheating the state, and to prevent civil servants and the government from cheating

the people, some systems are there as a result of overeager people trying to care for you and me. Systems that need to be in place to secure the welfare state are necessary, but privacy could probably benefit from system design taking privacy into account. Other systems though are unnecessary and should never have been deployed. One example is the camera surveillance of public transportation. A lot of research on CCTV and its effectiveness exists, most of it concludes that there is no significant reductions in violent crime in areas where CCTV has been deployed [15, 16]. The arguments used by the owners when they sought to deploy the surveillance of busses and trains, was that the surveillance was supposed to prevent robbery and violent episodes targeted at their employees [17]. While they are presenting it to the public as something they do for the safety of travellers.

According to [18], 91% of those asked are positive to surveillance of public areas, 79% states that cameras make them feel safer, but still 41% do not like the increased amount of cameras. One point in this survey that was particularly interesting was the fact that people who generally felt safe where they travelled daily, felt more safe if they were in areas with CCTV. People who answered that they felt unsafe did not report feeling safer in areas with CCTV. So it seems that cameras make people who already feel safe, feel safer. While people who does not feel safe are unaffected whether an area is under surveillance or not, aligning with research stating that surveillance does not significantly reduce violent crimes.

Another example more related to computers is the recent deployment of electronic tickets in Oslo, which will be further discussed in Chapter 3, where information about every trip made using a personal card is stored in a central database to provide end-users with different value-added services. But is this really something the end-users want? Some of them might value such services and are willing to trade in their privacy while others are not. In this, as in many other cases, the system owner decides what is best for their customers, and deploy an infrastructure in order to help them. However such infrastructures, as is the case with the one deployed in Oslo, might also double as surveillance systems.

To conclude, some of the systems eroding privacy are deployed as a result of a real necessity. However, many of these systems could have less impact on privacy if a more appropriate system design had been used. Some systems are deployed from a genuine wish to protect and care for customers and citizens, but some of these systems may have a high price in terms of privacy and society would be better of without them. While others again are built by small organizations or companies in order to protect their own assets. This might not have large consequences when done on a small scale, but when everyone implements such systems the total impact may be quite large.

2.4 Privacy and Personal Information

Up until now this chapter has been about what privacy is, and arguments for and against privacy in different settings. Hopefully the reader is now ready to dig deeper into how information is used in computer systems and how privacy is eroded in many cases. We start by defining some important terminology and concepts.

2.4.1 Personal Information

Personal information can loosely be defined as any kind of information about a person; ranging from his or hers favorite chocolate, to social security numbers and fingerprints. One thing to note is that it is not necessary for a piece of information to be directly linkable to an individual for it to be considered personal information. For instance by themselves the following pieces of personal information "24 years," "Computer Science Student," "from Fusa, Norway" do not relate to the author of this thesis. But if one knows that they are all information about the same person they might identify the author uniquely. Information derived from such facts is also in many cases considered to be personal information. Information about an individual's net income, mortgage, and payment history can be used to describe the person's financial situation from which one might further deduce whether or not he or she is to be considered a prompt payer. The European Data Protection Directive from 1995 uses the term "Personal Data" and gives it the following definition:

> *"'personal data' shall mean any information relating to an identified or identifiable natural person(data subject); an identifiable person is one who can be identified, directly or indirectly, in particular by reference to an identification number or to one or more factors specific to his physical, physiological, mental, economic, cultural, or social identity."*
> *[19, Article 2.a]*

2.4.2 Sensitive Personal Information

In many countries, a subset of personal information is given special protection, because it contains information considered to be sensitive. In Norway, as in most other EU countries, information about race, ethnicity, religious views, criminal records, health information, sexual preferences, relations, and union memberships are granted special protection under the law [20, §2 Sec. 8] [19].

However, it is the author's view that sensitive information should be thought of as "information an individual believes should be kept private," acknowledging that this is a matter of personal preferences. According to the survey mentioned earlier in the chapter, people's view of what they consider as personal information does not align with the current legislation [18]. In Norway people seem to be more willing to share information about their political and religious views and union membership, than their National Birth Number (NBN) and their cellphone number. Neither of these two numbers are granted any special protection by law, and the phone number is in most cases available in the phone book, and on numerous of online services. The authors of the survey point out that they believe people may have misinterpreted the question, and this possibility should be taken into consideration when trying to interpret the results. Another possible explanation is that how people may misuse a phone number or an NBN is simpler to imagine than how information about ones religion might be exploited.

Either way, the reader should now know that sensitive personal information exists, and in many countries such information is granted special juridical protection. One does not need to look more than seventy years back in time to imagine what consequences a comprehensive list of people's religious views would have had. So when developing a digital system that manages personal information, one should check whether or not the system will be dealing with sensitive personal information. If that is the case, one should do a thorough analysis of the privacy in the system design.

2.4.3 Identities and Identifiers

In different settings on the Internet we use different identities. On Facebook you present yourself with your given name, and an email, but on other sites people use pseudonyms, e.g. "ladiesman217." On government sites or in online banks, at least in Norway, the de facto standard to identify yourself is to use your NBN. All these identities contain a subset of an individual's personal information, and as we have demonstrated a person can have multiple identities, perhaps even overlapping identities. For instance an identity in a bank contains a lot of financial data that is mapped to an individual using an NBN, while the census authority has information about where individuals live, but also use NBNs to map an address to a unique individual.

When discussing identifiers, we see that some identifiers are in a sense "stronger" that others. A person's NBN is handed out at birth, follows the individual from cradle to grave, and never changes, except in some very special situations (e.g. sex change operations). In general an identifier is *strong* if it allows a unique mapping to a specific individual in a population.

Later in the thesis we will look at how broad, or even wrong, use of strong identifiers can reduce privacy. An alternative to strong identifiers is to use multiple identifiers that in themselves do not allow a unique mapping, but when held together provide a mapping with a certain degree of confidence.

2.4.4 Personal Information and Crime

The advent of the information age has undoubtedly made a lot of things easier for everybody, including criminals and fraudsters. The Internet allows for people located in one country to carry out certain types of crime on the other side of the world. According to Symantec's annual Internet threat rapport, the Internet has facilitated an entire underground economy where one can buy or sell information that can be used to commit different types of online fraud and other types of online crime [21].

One of the rising forms of crime seems to be identity theft. Identity theft occurs when someone uses another individual's personal information to impersonate him or her [22, p. 96]. Usually, identity theft is carried out with the intent of gaining access to some form of good, this can be everything from getting issued a credit card to using someones' store discount. When trying to commit crimes such as identity theft or fraud, having as much information as possible about the subject you try to impersonate is very helpful.

Identity theft combined with a general lack of security around personal information enables criminals to harvest vast amounts of information without making a lot of effort. A good example of harvesting is illustrated in [23]. We are shown how Norwegian telephone companies made very user-friendly sign-up procedures where all you needed to start the registration process was an NBN, filling the NBN into a form and pressing enter fetched your name, address, performed a credit check, and presented it all neatly on-screen.

NBNs are highly structured and therefore it is simple to generate valid numbers for any given day [24]. So by generating valid NBNs, and inputting them to the websites in an automated way, one could harvest a lot of information about Norwegian citizens. Namely the individual's NBN, his name, address, and an indication on whether or not he or she was to be considered creditworthy by the telephone company. It is important to note that one did not need to have a prior relationship with one of the the companies in order for them to leak your name, address, and NBN because the companies bought services from others who had more or less direct access to the census database.

A person's NBN was never intended as a way of authenticating that someone is indeed who he or she claims to be, but still many vendors used to grant access to various services based on the knowledge of an NBN. One

vendor went as far as issuing you a credit card as long as you had a valid NBN and a working e-mail address. In another recent article [25], a reporter investigated how much personal information about himself was lying around on the Internet. He found it to be frightfully much, the list included his Social Security Number, addresses dating back to 1975, and affiliations with various nonprofit organizations.

It is clear that a lot of information about people is already freely available on the Internet, and much of the information one can find there is regarded as public and should not be of any concern. In fact much of the openness is a requirement for a well-functioning democratic society. But accidents keep happening, and from time to time, some company or government branch experience a security breach and a following leakage of sensitive personal information. One can name numerous examples where personal information considered private has been leaked either by accident or foul play. For instance, the loss of a harddrive in Britain containing personal information about 100 000 soldiers and 600 000 potential recruits, or the New Zeelander who bought an mp3 player containing documents about American soldiers, camps, and supply plans in Iraq. The interested reader might visit the American nonprofit organization "Privacy Rights Clearinghouse" for a comprehensive list of American data breaches [26].

2.5 State of the Union

Several big organizations have already implemented routines and business processes to ensure that old systems do not constitute serious privacy risks, and that new systems under development are as secure and privacy friendly as possible. The United Stated passed the "E-government Act of 2002" that was supposed to improve the management of and promote the use of electronic government services. The act also went far in recognizing that the growth of new digital techniques and devices might have severe impact on the privacy of people's personal information. As a result, the act required that in the future, government agencies must analyze the possible ramifications of privacy whenever designing a new system, or substantially revising old ones. Such a review is called a "Privacy Impact Assessment"(PIA) and is meant to cover the most important parts of how a system handles personal information.

One of the governmental bodies in the US that provides a lot of online material about how to conduct a PIA is the Department of Homeland Security (DHS). Based on their web pages, and a memorandum from the Office of Management and Budget, I will give a brief overview of the most important

parts in a PIA. For more thorough information the reader should visit the web pages of the DHS [27].

The DHS material consists of nine sections where the people that are carrying out the evaluation are asked to answer certain questions. They are also given some guidelines as to how detailed the answers should be, and what they should cover. Each of these sections cover different sides of how personal information is managed, starting with questions elaborating on the management of the information, and ending with questions where the writers are asked to discuss how the subject at hand affects overall privacy.

The most important sections in a PIA are those that cover characterization and use of the information, retention and sharing. Also included in the PIA are sections for covering how those whose information is registered in the system are informed about it, and their rights to access, redress, and correct the information.

As we shall see in the next chapter, the division into these nine topics make very good sense when trying to determine, or document, the level of privacy in a computer system that processes personal information. In the next chapter I will introduce a number of controls that we will use to evaluate the privacy in a system. The general idea is that by ensuring that a few of these controls are in place we are able to say something about the overall level of privacy offered by a specific system. The reader will then see how subjects included in the DHS PIA falls in under specific controls.

Chapter 3

Measuring Privacy in Computer Systems

Quis custodiet custoder ipsos?—Who watches the Watchers?

The previous chapter gave an introduction to privacy, and looked at how a few organizations and government bodies are trying to ensure that their information systems preserve privacy. This chapter starts with an introduction to risk management in computer systems, how it is done, and some remarks about what is considered best practises. In risk management there are two main methods used to quantify risks, namely qualitative and quantitative risk management, both methods will be introduced and it is outlined how they are used to measure risks. The risk management introduction is followed by the introduction of privacy controls. I suggest some controls that will aid in the process of evaluating the overall privacy level offered by an information system. These controls will then be used at the end of the chapter, as parts in a proposed formal privacy reviewing process inspired by risk management.

3.1 Risk Management in Computer Systems

Risk management is a process that enables those performing it to identify the risks that they, and other stakeholders are exposed to and manage these risks. The goal is to determine which, if any, of the identified risks that are too high, and to devise strategies to remove, or at least, mitigate these risks. The National Institute of Standards defines risk management as *'the process of identifying risk, assessing risk, and taking steps to reduce risk to an acceptable level"* [28].

Risk is defined differently depending on the setting. For instance, when playing the lotteries you can either lose or win. Since playing the lottery usually involves placing a bet there is a risk involved, this type of risk is often referred to as speculative risk. Risk management focuses solely on managing risks that have a negative impact on a business or a system. This type of risk is referred to as pure, or non-speculative risk. So to us a risk is simply the possibility of suffering loss or harm.

When developing software it is usually a requirement that the software is secure, meaning that it should be able to function correctly under malicious attacks [29]. Many different methodologies for development of such software exist. Common for most of them is the use of a risk management process, that is to help reduce the overall risk level of the application. Most such development methodologies consider security, but some of them focus on security more than others, amongst these are Microsoft's Secure Development Lifecycle (SDL) [30]. However, the different methodologies incorporate a lot of the same "best practices" in order to achieve a high level of security, and these are summarized in Figure 3.1. In this Figure, a general Software Development Lifecycle(SDLC) is seen with the different best practices placed where they generally belong; showing that risk management has grown to be a well established practice in software development, and that it should be considered an important contribution to software security.

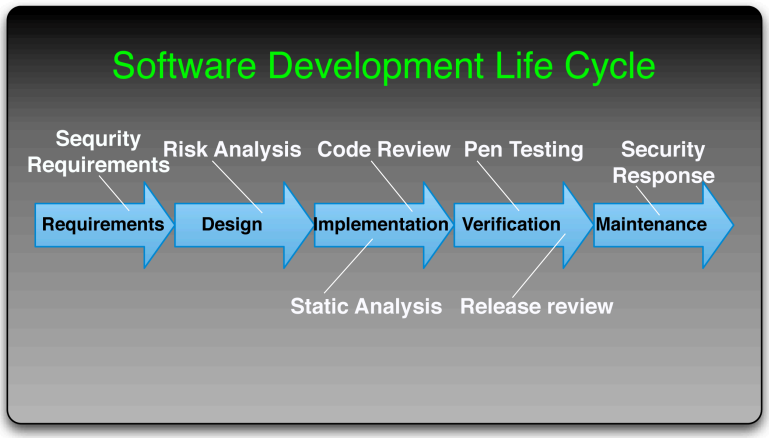


Figure 3.1: The SDLC with all best-practise security measures filled in, the figure is inspired by [31].

In general, a risk management process consists of two steps, risk assessment followed by a risk treatment. Figure 3.2 gives a high level view of the process [28].

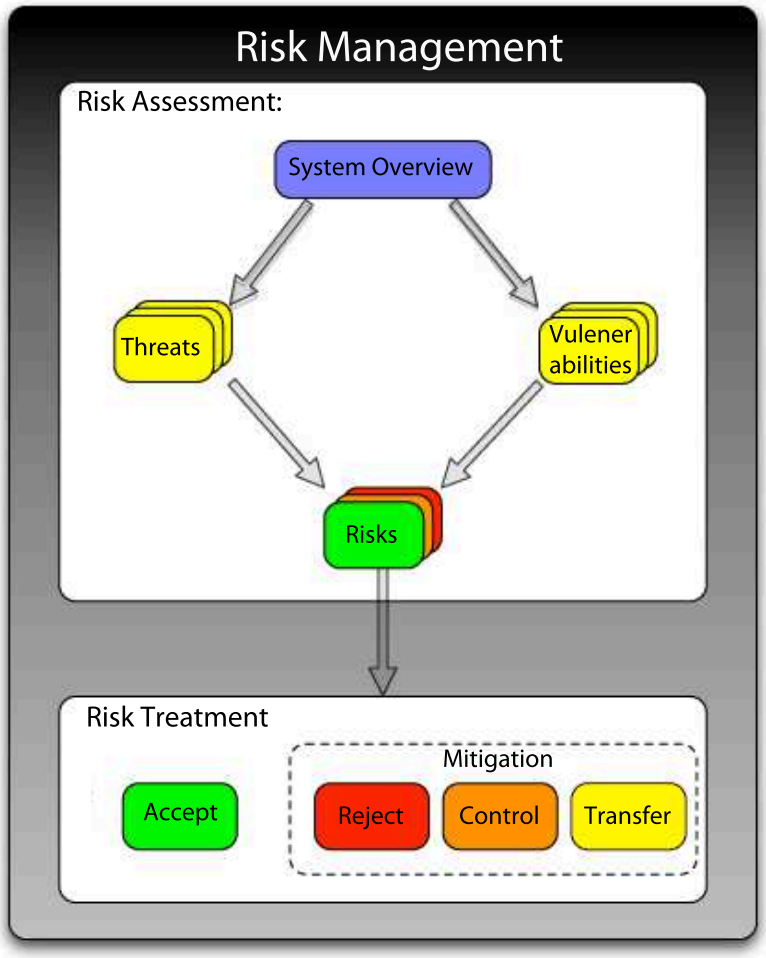


Figure 3.2: A high-level view of a risk management process.

3.1.1 Risk Assessment

The desired output from a risk assessment is a set of quantified risks. In order to achieve this, one starts off by getting a complete system overview, identifying assets, stakeholders, and describing the purpose of the system. An asset can be everything form a physical hardware device to a piece of software or data. The importance of data that resides on an organization's information systems should not be downplayed. In many instances the data are far more worth than both the hardware and software. Identifying the stakeholders of the system is also important, because we should consider risks affecting other stakeholders than the one carrying out the risk management process. There are several reasons for considering risk to other stakeholders. If shortcomings in your system inflicts damage on someone else, do you have

to compensate them? Also, incidents not harming you directly can do so indirectly, for instance from to bad publicity leading to loss of reputation.

After mapping out the system, the two next stages in the risk assessment is threat and vulnerability identification. A *threat*, or a threat agent, is defined as an entity that may cause harm to your system by exploiting a vulnerability, either intentionally or unintentionally. Threats can be hackers, script kiddies, insiders with different types of privileges, users behaving badly, and natural disasters such as fires and earthquakes. A *vulnerability* can be said to be susceptibility to some kind of injury or damage, or a weakness in your system allowing someone to inflict damage by exercising the weakness.

Having identified vulnerabilities and threats, one goes on to combine these into pairs of threats and vulnerabilities. Afterwards each of these pairs are given a likelihood, indicating how likely it is for the threat to exploit the vulnerability.

To qualify as a risk there has to exist both a threat and a vulnerability. If there is no-one around to exploit the vulnerabilities, then they do not constitute risks. Likewise if there are no vulnerabilities the number of threat agents does not matter.

Having paired off threats and vulnerabilities one continues to sort these based on how severe an impact they can have on your business. How the ranking of risks should be done will be further discussed in Section 3.1.3

3.1.2 Risk Treatment

After identifying and ranking the risks one has to decide how to manage the individual risks. Some of the risks may be negligible and can be ignored. Ignoring a specific risk should of course always be a result of careful consideration and generally a cost/benefit analysis is used to make the decision. The acceptance of a risk can happen when the impact of an incident is so small that just dealing with incidents as they appear consume less resources than fixing the problems. Some of the risks may have a low probability of occurrence, thus one can assume that they occur so seldom that they too would cost more to prevent than to accept. Deciding to not mitigate risk is referred to as *accepting* the risk.

Other risks can be so severe that one has to do something to *mitigate* them. Different ways of mitigating severe risks exist. For instance one might be able to *transfer* the risk. Transferring a risk is, as the name implies, to take a risk that you are exposed to and somehow transfer it to other stakeholders. Risk transfer is typically done every day by car owners, when they buy an insurance they transfer most of the risk in the case of an accident to an insurance company in exchange for a fee. One can also choose to reduce the

risk by implementing different kind of controls and safeguards. In banking, an example of a safeguard is the requirement of having all transactions above a certain value authorized by two clerks. Introducing different measures, or safeguards, in order to reduce or eliminate a risk is called *controlling* the risk. Another option is to *reject* the risk, and choose not to offer a specific functionality in the system because the risk it would introduce is too high. A simple illustration can be seen in Figure 3.2 where the risks identified in the "Risk Assessment" step are distributed into different categories.

3.1.3 Quantifying Risk

What risk mitigation strategy to choose depends on the severeness of the particular risk. Thus, how risks are quantified is crucial and should reflect the world as accurately as possible. Focusing the effort and attention where it is most needed, gives the best return of investments, and reduces the overall risk exposure as much as possible, hopefully keeping incidents under control.

As mentioned at the beginning of the chapter, there are two main methods used when quantifying risks. When using a *quantitative* risk management approach one assumes that over time the incidents have a statistical distribution, e.g. a normal distribution, and one can do statistical computations on the probabilities. In such an approach the threat/vulnerability pairs are given estimates of impact, usually in monetary loss, and a probability of occurrence. Then the risk is calculated using the following formula [32]:

$$Risk = Impact * Probability\ of\ occurrence \tag{3.1}$$

With such an approach one is able to calculate the expected cost of a risk. If for instance a certain type of incident would cost a company about 1 000 000 NOK but only have a probability of 0.5%, the risk could be said to be 5000 NOK, which is the expected average loss of that specific risk. For another similar incident were the probability is 5%, the risk would be 50 000. Here, we would start by focusing on the latter risk, as it is the one we expect to cost us the most.

In some areas the quantitative method might be applicable and produce sensible results, while in others it might not yield good results. The underlying assumption of quantitative risk management is that considering probability of occurrence makes sense, and that by looking at the past one is able to predict the future. Even when dealing with systems where the assumption is true, one need to have enough data points for these probabilities to be accurate. As most companies does not want to talk about their security

breaches and incidents, compiling a high quality list of data points might be very difficult.

And even if one manages to create a perfect estimation of the probabilities, the method is still subjective since the monetary loss of incidents is hard to predict and can vary over time. For instance, downtime in an online bank might be cheap during the weekend, but what if it happens the week before Christmas? For those who are interested in more criticism of quantitative risk management, or risk management in general see Taleb [33].

The other approach to risk quantification is *qualitative* risk management. Using a qualitative approach to risk management one decides on a set of discrete levels that are used to classify the likelihood and the impact of a risk. A common choice for the levels is "high," "medium," and "low" for both impact and probability, and then one combines the two to create the final level for the risk. The risk can then be determined using a risk matrix, as shown in Figure 3.3. The qualitative approach also has its weaknesses. To a certain degree, it underestimates the risk associated with high impact, low probability events. As we see from the risk matrix, a high-impact, low-probability event ends up classified as a medium risk, but such events can be catastrophic for a business. This is one of the major weaknesses Taleb points out in his book [33]. In particular, the observation holds true for large national or international systems of great importance. A discussion of how to manage these risks is found in [34].

Nevertheless, I argue that the qualitative approach is the best alternative when securing most computer systems, and the method in [34] only comes into play in specific cases when talking about large systems of national or international importance. Using qualitative risk management, enables risk managers to focus on managing the different risks, and less time worrying about whether the values for impact and probability are correct. As the saying goes "it is better to be approximately right than exactly wrong."

3.2 Privacy Controls

In the previous section, we saw how a qualitative approach is valuable in risk management of computers, and how it is a tool highlighting the most critical problems of the system under review. Inspired by the qualitative approach to risk management, we now turn to analyzing privacy in information systems.

In a system maintaining personal information, the approach of identifying vulnerabilities and threats can be a viable approach to determine the level of privacy. But risk management is not about creating 100% secure systems,

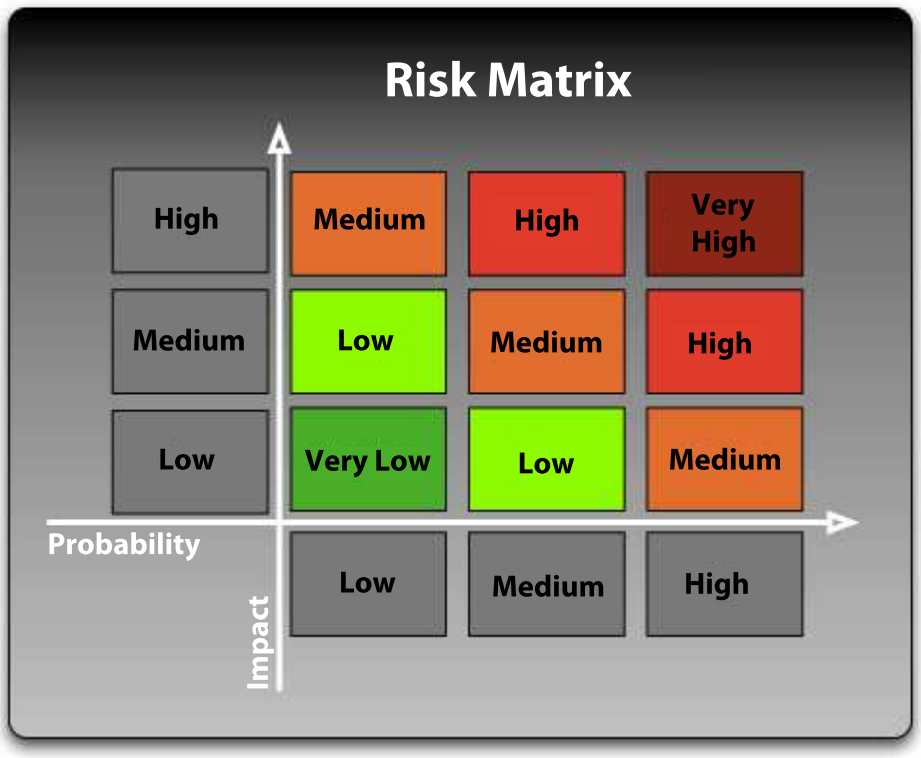


Figure 3.3: An example qualititative risk matrix.

it is about creating systems that one can afford to live with and planning for the unexpected. When dealing with people's personal information, such a gamble may not be desirable. Therefore, I believe that a different approach should be used when trying to ensure that a system provides privacy.

The rest of the Chapter is devoted to defining nine general areas where privacy fails in computer systems. For each area, I introduce the notion of a *control*, that should be understood as the ability to manage the privacy issues related to the specific area. Using this approach one should be able to identify privacy problems in systems, because one has a clearly defined standard that one can measure against. The approach also enables organizations and individuals to take a structured approach when trying to evaluate the privacy of any given system.

Through the work with my thesis I have found if useful to define two different classes of systems providing privacy namely, systems that offer their users *anonymity* and systems that offer users *privacy of their personal information.* I will introduce the separate controls for each class of systems. The word "anonymity" is derived from a Greek word that originally meant *without a name*, and in a digital world we should think of a subject as anonymous if it is not identifiable within a specific population. While privacy of personal information follows the definitions that were given in Chapter 2.

3.2.1 Privacy Controls in Systems Offering Anonymity

Many systems do not need their users or data subjects to uniquely identify themselves in order to deliver services. Such systems span from simple webpages that just provide readers with static information to large multi-user systems that process information for various reasons. There seems to be a lot of scepticism towards offering true anonymous services though, largely due to the potential for misuse. The upside of anonymous services however might be greater than the downside, and in some cases the ability to be anonymous is a life and death issue. For instance many human rights organizations rely on anonymity services in order to stay in contact with people in countries where freedom of speech is limited. A negative comment about the Chinese government can in some cases be enough to land you in jail, and it has been known that Chinese human rights fighters use The Onion Router (TOR) network to communicate with the outside world. The TOR network facilitates anonymous internet access originally developed with the support of the US Navy [35].

Recently there has been some outcry in Norway due to a new electronic ticket system in Oslo. Users are complaining that the system is logging far too much information about them, e.g. when they travel, where they

travel from, and what direction the vehicle they entered is traveling. All the information collected by the system is available to everyone with access to the system, and the user's full name, address, and birth-date are stored in the system [36]. A thorough analyze of how such a system could be built to offer anonymous traveling is outside the scope of this thesis, but is it really necessary for every company in such an electronic-ticket collaboration to have access to your full travel history? I for one think not. Fare collection systems for public transportation should be built in accordance with the right an individual has to free movement, something that one can argue a thorough logging of ones movements is not. The wish to store as much information as possible seems to be somewhat symptomatic for new systems that are being deployed.

In order to offer anonymity, a system needs to have two properties. One should not be able to trace an action performed in the system back to a user, and one should not be able to deduce which actions in the system are performed by the same user. These two properties will be referred to as *untractability* and *unlinkability*, and will be closer described in the two following sections. Both properties are defined as controls in systems offering anonymity.

Untraceability

It should no be possible to trace actions back to a user.

As stated, a system that aims to offer its users anonymity should be built in such a way that actions performed in the system cannot be linked to a specific user. Traditional election schemes require that it should not be possible for anyone to determine what a specific individual voted for, while you usually have to register yourself in order to place a vote. Voting systems are thus good examples of systems that contain personal information, but still ensure that an action, in this case posting a vote, cannot be linked to the individual who performed it.

On the Internet, the use of pseudonyms is a common way to obtain a certain degree of anonymity, since a pseudonym generally is not easy linkable with an individual's real identity. However, when using the Internet, if no precautionary steps have been taken, every service one visits know the originating IP address, and thus users are largely traceable.

Generally, in any system, the harder it is to track a specific task to a unique individual the higher the level of anonymity is expected to be.

Unlinkability

It should not be possible to link actions performed by a specific user.

If a specific user interacts with a system that is to offer anonymity multiple times, it should not be possible for anyone to create a list containing the different tasks performed by a single user. One should note that the ability to generate such a list does not translate into determining who the user is. Just that the same user performed all the tasks in a list. Over time, a list can grow quite extensive and it might be possible to use bits of information from different tasks, viewing them together in a wider context, to uniquely determine the identity of the individual behind the actions.

In 2006, an employee at America Online (AOL) published search queries for 650 000 AOL users from within a three month period. The data was anonymized before they were published by replacing the AOL username with a serialnumber. By doing so they thought that researchers could use the dataset without interfering with the privacy of those who had done the searches. AOL was wrong. By looking at the content in the search strings it was possible to deduce the person behind the number. Something the New York Times decided to do with user 4417749. User 4417749 had done several hundred searches in the three month period for which data was published. Some of the queries used to track down the person hiding behind the number are listed below.

- 60 single men.
- Numb fingers.
- Several queries involving people with Arnold as their last name.
- Landscapers in Lillburn, GA.

From these strings one can assume that the person is a female, living somewhere in Lillburn Georgia, she is probably in her 60s and single. The reporter from the Times did not user very much time to track down 62 year old Thelma Arnold from Lillburn Georgia, a widow with a dog and an interest in her friends ailments [37]. A good illustration on how different pieces of information, that by themselves do not necessarily identify an individual, when linked together can reveal a lot more information than what was originally intended.

3.2.2 Privacy Controls in Personal Information Systems

While the two previous controls address key issues related to providing anonymity in a general information system, the controls introduced here will be of use in systems where the goal is to keep users' private information confined within the specified boundaries. In real world applications, the nature of personal information being stored is diverse, and for some sensitive types of information, a privacy breach may cause permanent damage to the individuals whose information is disclosed. A typical example of such data is health-related information as there are many different diagnoses in the world, some with greater stigma than others.

The fact that a system needs sensitive personal information in itself should not disqualify it from being built, but care has to be taken to build in privacy and security from the beginning. As with security, privacy is generally not something that can be added to a system once it is up and running, it has to be introduced from the very start of the development process.

Collection

Any information system should only collect the minimum amount of personal information that it needs to fulfil its purpose.

By collecting more data than really needed the potential loss in the event of a data breach is higher that necessary, something that may have unfortunate consequences for your data subjects, whom in many cases are your customers. Doing so may also make the system seem more invasive than it really is, as the list of information it collects is unnecessary large.

One cause of privacy erosion is *function creep*. Systems that are built for one purpose collects, processes, and stores information to solve a specific problem. Over time other problems that might be easily solved using the same information arise. Since the information is already collected, it is easy to use it in order to solve the new problem. For instance, in Norway we have recently had a spike in the number of roads that require toll, and an automatic system for tolling named autoPASS has been introduced in major cities. At the start of 2009, thirteen of the toll-financed roads in Norway used autoPASS to do road tolling, including the city centers of Oslo and Bergen. In order to carry out the automatic tolling, the autoPASS system registers information every time a car passes through the toll-gates, and stores the information in a central database. After a couple of years of using the system, tax authorities in Norway realized that they could use information from it to determine whether or not people were paying the taxes they should for

private use of company issued cars. If the tax authorities wants to build such an infrastructure themselves, it would lead to public outcry and it would never happen. But by using registers already available, the public outcry may not be large enough to prevent them access to the register. At the time of writing, it has not yet been decided whether or not the tax authorities should be granted access to these data.

The purpose of systems that have an impact on the users' privacy is important when deciding whether or not it should be built. For some systems, the advantages may outweigh the disadvantages. Automatic tolling systems is a good example of such a system. On one hand it allows for a better traffic flow, and is very convenient for all parts. On the other hand the system deployed in Norway today have a huge impact on privacy by storing time, date, and position for every passing of a toll gate. Most people are willing to compromise and trade some of their privacy for convenience. But for those that are not, the system should have alternatives, which is not the case today.

Once the system is in place, like with AutoPASS, other ways to use the information is thought of and implemented. The new uses of the information may be far more controversial than the initial use, and had it been known beforehand it might have lead to the system's dismissal. To combat function creep as much as possible, systems should be designed to collect as little information as possible.

Retention

Personal information should be retained for the shortest possible period.

The shorter the period of retention, the higher one can expect the level of privacy to be. In a system where personal information is deleted after a short time, there is, naturally, less information available in the case of a breach. Short retention periods also increase the level of privacy by reducing the historical data available to system owners.

When discussing retention of data, a brief discussion of how such historical data might be used, and or misused, is in place. The last decennium different knowledge discovery techniques have gained a lot of steam, and especially data mining and automatic profiling. With the advent of the information society, the amounts of data available have grown to the skies. To make sense of the incredible amounts of information one need ways of sorting the information and finding those pieces of information that are truly interesting. The aforementioned knowledge discovery techniques attempt to solve the problem.

Automatic profiling is large research topic by itself, but it is basically about dividing pieces of information into classes, based on predefined prop-

erties. Generally, in order to do profiling one has to have a purpose for the profiling, one needs to specify the different profiles, and apply them to a data set. By doing this one can identify those types of information that are particularly interesting and discard the rest.

Such techniques are used in several areas, for instance in finance to discover fraud by looking for unusual transactions. Another area where such technologies are used is credit rating. As an example, one might have a class for prompt payers, one for those who are slightly more slack, and one for those who time after time fail to pay bills before they are due.

Dependent on how such techniques are applied, and their quality, such data mining may have unwanted effects. How would you like to be wrongly classified as a bad payer, or be placed on a no-flight list because automated processing says you are not behaving like "normal." In a recent book about data mining and profiling, Brownsword discusses what the consequences of a society pervaded by profiling might be [38]. Portraying some of the same dangers as Gaycken [12], fearing that it may impact peoples ability to make ethical choices.

As with many other surveillance and privacy invasive techniques and tools, several governments have been experimenting with automatic profiling of passenger data on international flights. Whether the tests have been successful or not is unclear, but a recent report indicates that it was not as fruitful as hoped for [39].

Secondary Use

Collected personal information should only be used for the specific purpose it was originally collected.

As mentioned in the paragraph about collection, function creep is a large problem. In many cases only collecting the minimum of information needed might not be enough. The information might still be valuable for someone else in order to solve a different problem. Again careful design and implementation will probably get organizations some way in ensuring that the information may not be used for other purposes than originally intended. But steps to prevent secondary use do not need to be of a technical nature. For instance, having good contracts with users regulating how the information gathered by the system should be used, can also help.

The new fare collection system in Oslo mentioned earlier seems to store far more information about a user than what should be necessary. The reasons for storing all the information are to avoid disputes about the amount of money left on the cards, and to be able to re-issue cards in case one is lost. But if an insider knows your name, he or she is able to get a complete list of

your recent travels, something that is not very privacy enhancing.

Distribution

The personal information collected by any system should not be made available to third parties without prior consent from the data subject.

If a system enforces this control, it assures the data subjects that their information will not be shared with third parties, and thus enable people to have a certain degree of control over who has access to their records. For anyone to claim that their system offers a high degree of privacy, it has to employ measures to prevent the distribution of personal information. In Chapter 2, we saw that most definitions regarding information privacy one way or another included the ability to selectively disclose information about oneself to others. If an organization collects personal information about people and willingly shares it with others, either free or as a paid service, they rob the data subjects of their right to privacy.

Distortion

Operators of a system should do their best to assure the integrity of the information system.

Steps should be taken to ensure that the data is protected against different forms of distortion, both unintentional changes stemming from errors, and malicious attacks to change the information. At the very least, such changes should not be allowed to occur without detection.

An example of how dangerous such distortion can be was illustrated during a red team exercise in 1998. In computer security, a common way of doing penetration testing of systems is to use a red team, blue team approach. The blue team is supplied with information about the system while the red team is not, and they are then both tasked with compromising the system [40]. During the exercise in 98, the red team was able to compromise a Department Of Defense (DoD) website containing personnel records and found themselves able to alter the blood type of soldiers. Although the system they exploited was merely a demonstration system, the potential of distortion in mission-critical systems, whether deliberate or erroneous, should be clear [41, 42].

Correction

Any individual whom the system stores personal information about should be able to access and correct data concerning self.

For any system, a good description of what information the system collects should be available, but it should also be possible for users to see what information the system has about themselves, and if the information is wrong correct it. So, for systems maintaining personal information there should be routines in place ensuring that such actions are possible. In Norway, companies that maintain personal information are obliged to provide the following information upon inquiry [20, §18]:

- Name and address of those responsible for storing and maintaining the information.
- Why the information is being gathered.
- A description of the information that is used.
- Where the information was collected from.
- Whether or not the information will be shared and to whom.
- If you are registered, you are entitled to know what data is registered about you.
- Which safeguards are in place, as long as publicizing information about the safeguards does not reduce the security.

In addition, the same law includes a paragraph requiring those that maintain personal information to correct obvious errors, as well as errors pointed out by those registered [20, §27]. Care has to be taken as wrongly corrected information can also have negative consequences.

Notification

In the case of a mishap the users whose personal information was leaked and, perhaps, misused should be informed of the incident.

When a company has experienced a breach in a system, it should notify the users potentially affected by the breach. The reason for giving such notification is simply that if someone looses your information you should be aware of it. Being aware that your personal information has been lost enables you to prepare for consequences that might follow from such a breach. For instance, if the intruders get hold of information that enables them to carry out some kind of identity theft, you can be on the alert and take some precautions. Depending on where in the world you live there are some steps that can be taken to reduce this risk. In Norway, it is possible to instruct the

credit rating companies not to give credit-ratings unless they are supplied a specific password.

Although giving notice can probably be considered as the right thing to do, many companies are reluctant to do so. The main reason for this is probably all the bad press that follows from it. Recently, the American company Heartland, an electronic payment processing company, went public with a breach that happened at the end of 2008. At least they went public, but they choose to do it on the day of president Obama's inauguration, and one can only speculate whether it was a strategic move to avoid media attention.

After 2003, several states in the US have introduced disclosure laws that require businesses to report breaches in a timely manner if personal information has been lost, is expected to have been lost, or acquired by unauthorized agents. Some states even require that those affected is to be addressed in writing.

3.3 Privacy Management

With inspiration from risk management and the various Privacy Impact Assessment examples mentioned in Chapter 2, the rest of this Chapter will give the reader an overview of how to use the proposed controls to evaluate the overall informational privacy offered by a specific system.

A general privacy review of a system will start by getting a good overview of the system in order to document how it is intended to work, or how it works. It is especially important to get a full description of how data flows around in the system, mapping out how it is collected, how it is stored, what it is used for and so on, describing the full life cycle of personal information in the system.

When a complete overview of the system has been acquired, the next step is to determine which of the controls introduced in Section 3.2 are applicable to the system. One can imagine cases where some of the controls do not come into use, for instance in systems where one has to keep people accountable for their actions it may not make sense to talk about the two controls regarding anonymity. The next step would then be to walk through the relevant controls, giving a thorough description of the relevant system parts for each of them and ending with a summary.

In the following sections we will go through all the steps and give a brief overview of what they should contain.

3.4 System Overview

When performing a privacy analysis of a system there are many interests and boundary conditions that have to be considered. So, getting a good overview of the system, what is actually does, what it is intended to do, and why, are all important.

Since we are mainly interested in personal information, the overview of the system should focus on how such information is treated. I earlier mentioned the information life cycle, and using this as a starting point is a good idea. Figure 3.4 an illustration of how one might think of the life of pieces of data. This figure shows that the initial data is collected from some source, then goes through an initial processing, before it is transfered to storage where it resides until it is used. In some cases the information in a system might be shared internally with other systems within the organization, or it might be shared externally for some reason. Finally, after a certain amount of time, the information may be deleted from the system.

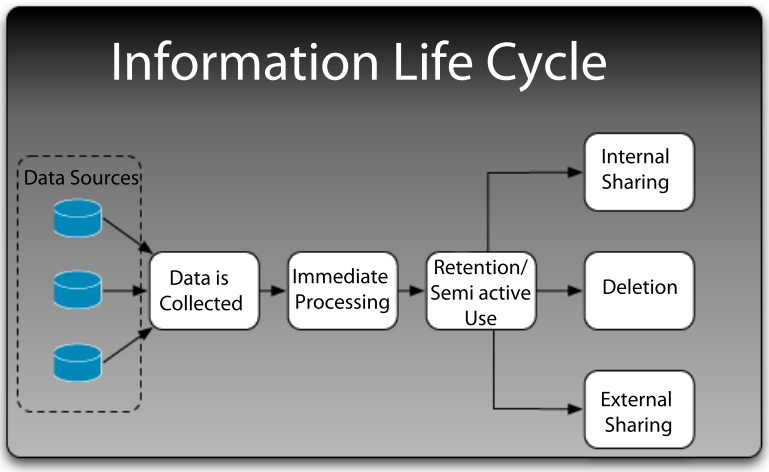


Figure 3.4: A illustration of how information generally flows through a system.

3.4.1 Metrics

The analysis attempts to measure the level of privacy in a system, thus some kind of metric is needed. Based on the earlier discussion of quantitative and qualitative risk management, I suggest that such a metric should be based on a one-dimensional High-Medium-Low quantification. Risk management

usually uses a 2-dimensional matrix as described in Section 3.1.3, and illustrated in Figure 3.3, this is due to the traditional view of risk as likelihood times impact. When evaluating privacy in computer systems, one tries to determine the level of privacy in the system and thus solely concentrates on the level of privacy and not the likelihood that privacy is breached.

It should be noted that the suggested privacy review somewhat overlaps with a risk management process, but they do not completely overlap and thus neither one should be seen as an replacement for the other. A risk management process focuses on preparing for incidents that may occur in the future and takes steps to ensure that the consequences of such incidents are as small as possible, while a privacy review focuses on determining how well privacy is taken care of in a system and highlights areas where problems reside. As a result I suggest that the individual controls should be given a High, Medium or Low rating dependent on how good the system protects privacy. An example of how the different levels should be used is given in Figure 3.5.

For the metric to be useful to anyone, care must be taken when the criterion for rating a control High, Medium, or Low are chosen. As the criteria for each level is set by those performing the process, it is possible to adjust them to make a system appear more privacy preserving than it really is. One way to counter this is to make the criteria public so that others may understand the assessments underlying the final privacy ratings. Personally, I believe that the high level should be considered as a nearly impossible goal, only to be used in cases where the system does a very good job of protecting people's privacy.

3.5 Analysis of Controls

Having completed the system overview, one should start analyzing the controls. For each of the controls one should start by determining if the control is applicable for the system in question or not, and if not the reason for this should be documented. Another special case may arise if those carrying out the review do not have enough information about a specific control. If there is no information the entire control should be marked as missing, stating that very little about it is known and that it should be rated low. If some information is available, it may be possible to carry out the evaluation, but the lack of information should be duly noted in the review. After completing the review of a particular control, one should try to determine the level of privacy offered from low to high. This involves setting some clear criterions for the different levels. This is needed on a case to case basis as every system

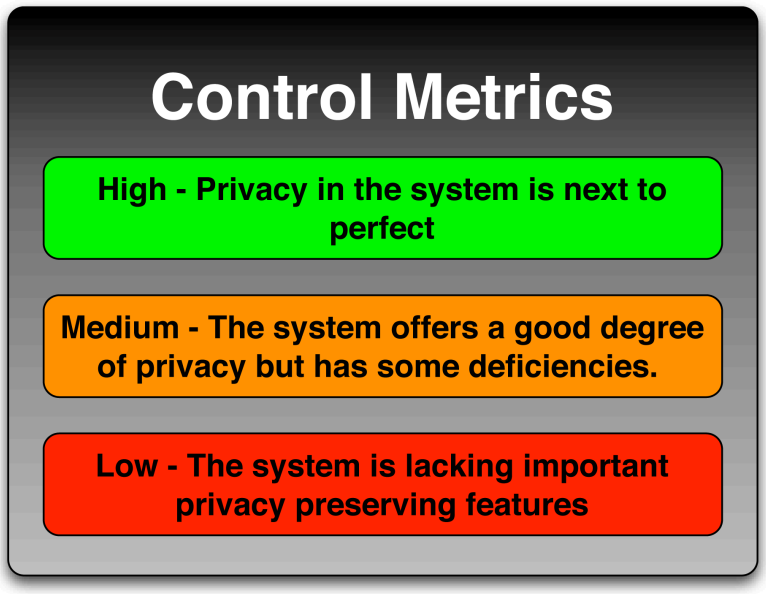


Figure 3.5: An illustration of the High, Medium and Low privacy levels.

is different, and to determine some generic requirements that can be used on every conceivable system is at the very best difficult, not to say impossible. Many of the issues addressed in the different controls are also subjected to legislation in some countries, and it would be beneficial to include a discussion of whether or not the system complies with this legislation.

In the end one should have a good system overview and a thorough explanation of how personal information and/or anonymity is treated throughout the system and an understanding as to what level the different controls are satisfied. This review should again enable the organization to prioritize the areas where the attention should be focused.

Chapter 4

Privacy Management Example

The poorest man may in his cottage bid defiance to all the force of the Crown. It may be frail; its roof may shake; the wind blow through it; the storms may enter; the rain may enter — but the King of England cannot enter; all his forces dare not cross the threshold of the ruined tenement!
Parliamentarian William Pitt during a speech in 1763

In the previous Chapters we have been given an introduction to privacy and a structured approach to reviewing the privacy of a system maintaining personal information was outlined. The purpose of this Chapter is to apply the method outlined to a real information system, as an example of how it could be done.

In Norway the government is soon to decide whether to implement the European Data Retention Directive or not. This directive mandates prolonged storage of traffic data, but not content. It is fairly simple to imagine how content of communication data might be used, or misused. But how traffic data might be used, and what it actually is, is somewhat more difficult to grasp.

As a result, I have chosen to look closer at the privacy of personal information in Norwegian cellphone systems, and especially on traffic data. One thing that is particularly interesting about such data in the cellphone system, is that they incorporate the geographic position of the mobile station at the time it receives or sends data.

The Chapter starts with an introduction to the most relevant entities in a generic cellphone system. It then continues to describe how geolocalisation of a cellphone can be carried out. This section of the chapter also contains information about a simple piece of software that I wrote to exemplify how

geolocalisation can be done. The last section contains the actual review of how personal information is handled. Readers should note that the review is largely carried out with information publicly available, and is done on a general level, as opposed to look at one particular provider. Although, under some of the controls, examples from different providers are given. It should again be stressed that this review is made with the information available to me at the time, and that if I had been given more information the rating might have been different.

4.1 Cellphone System Overview

The Global System for Mobile communications (GSM) is the dominant standard for cellphone communication in the world today, and the system used in Norway. GSM as a standard was developed during the 1980s, and the first GSM release came in 1991. Since 91 it has undergone many revisions, incorporating new functionality and improving old. When released, it was considered the second generation of cellphone systems, while all the older analogue systems were labelled first generation. After some years, with the growth of the Internet, the need for improved transfer speeds of data became clear. In order to increase the transfer rates, the General Radio Packet Service (GPRS) was developed for use with GSM and other similar 2G systems, as an add-on. The combination of GSM and GPRS was considered a stop-gap effort on the way to new 3G systems, and was labelled 2.5G.

The 3G system deployed in Norway today is the Universal Mobile Telecommunications System (UMTS). UMTS was derived from GSM with GPRS and is thus backwards compatible, allowing for a gradual replacement of the old network infrastructure. All these systems are introduced here because GSM is the standard that lies at the bottom of most cellphone networks today, and with UMTS under deployment in the most profitable areas. Figure 4.1 illustrates the key elements in a phone network, where GSM and UMTS are deployed alongside each other.

As a general rule, both systems can be divided into three similar subsystems, although they have different names. For GSM these are the Network Switching Subsystem (NSS), Base Station Subsystem (BSS), and User Equipment (UE). In UMTS the NSS equivalent is named Core Network (CN), UMTS Terrestrial Access Subsystem (UTRAN) is the UMTS version of the BSS, while the User Equipment (UE) is the same in both standards.

In the following, I give a brief description of the elements in these systems, as an understanding of how the systems work is needed to explain how personal information is generated and how it flows around in the systems.

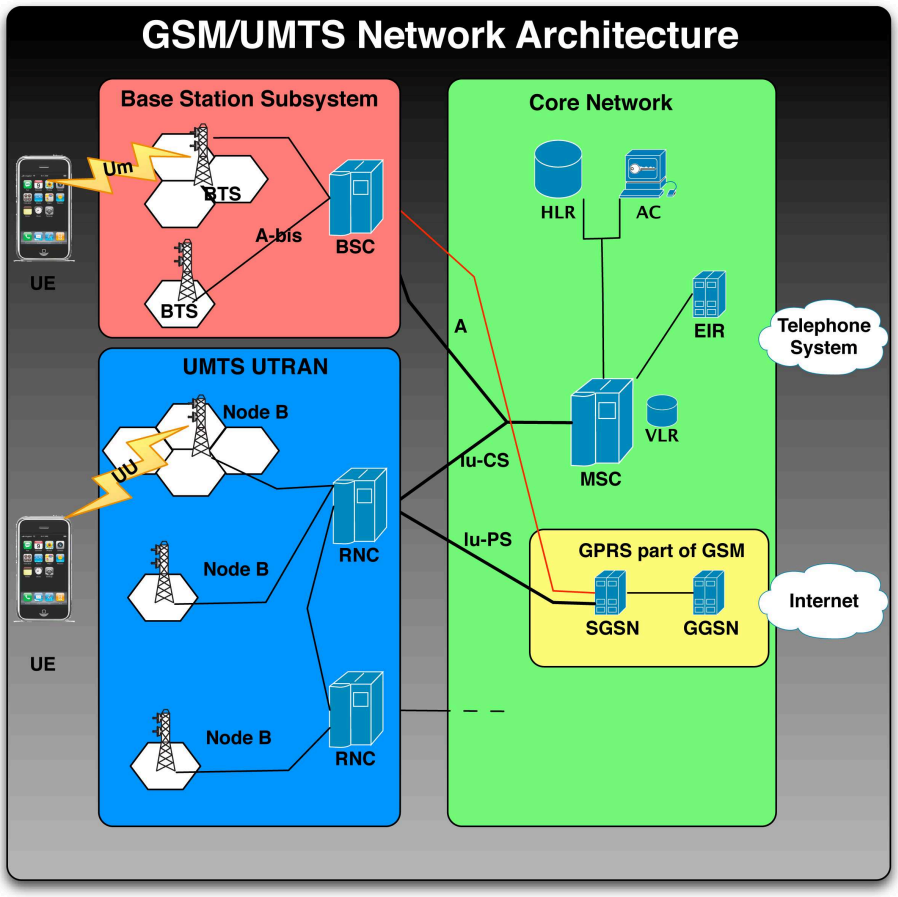


Figure 4.1: Illustration of GSM/UMTS architecture.

4.1.1 User Equipment

A UE is a device used by an individual to connect to the network. Usually this is a cellphone, but it may just as well be a dedicated modem used for connecting to the Internet via the phone network. Such devices are thought of as consisting of two parts, the first one is the handset, or modem, itself called a Mobile Station (MS) while the other one is a (Umts) Subscriber Identity Module ((U)SIM) card. USIM for UMTS, and SIM for GSM. The cards are the property of the network operators and contain the necessary information for subscribers to authenticate themselves to the network. Each card has a unique number, called International Mobile Subscriber Identity (IMSI), which is 15 digits long. In addition, the card contains an authentication key that it shares with the authentication center of the owning company. They also contain some other keys and numbers, as well as storage capacity for a

phonebook and Short Message Service (SMS) messages.

During the last two decades cellphones have evolved from simple phones to mobile computers with built-in phone functionality. Phones with functionality that resembles a home computer is referred to as a smartphone. In 2008, smartphones really hit the marked, Apple released their iPhone 3G, Google's mobile phone operating system Android hit the market, and other cellphone vendors released new "pro"-models. These phones often come with chips supporting a variety of technologies such as Bluetooth, Wireless Networking (WiFi), and the Global Positioning System(GPS).

4.1.2 Base Station Subsystem/UTRAN

In the GSM and UMTS standards, the part of the infrastructure responsible for handling signalling and traffic between UE and the telephone network are given different names. But both contain essentially the same equipment, although in different versions, and have similar responsibilities. In GSM the radio towers are called Base Transceiver Station (BTS) while in UMTS they are named node B, see Figure 4.2. Each tower is connected to a unit that manages one or more towers simultaneously, in GSM such units are named Base Station Controller (BSC) while in UMTS they are called Radio Network Controller (RNC). An illustration of the structure of these can also be found in Figure 4.2.

BTS/ node B

As stated, BTS and node B are basically radio towers with senders and receivers, providing the wireless link between the cabled telephone system and the UE. A tower is equipped with one or more directional antennas, each responsible for the signalling in one "cell," hence the name. In GSM the radio band is divided up into channels, and each such channel can only carry eight concurrent conversations. So, if an entire city was to be covered by just one cell, only eight conversations would be possible at any time. In order to combat this problem the size of cells are reduced, and the number of cells is increased, taking care that no neighboring cells are using the same radio frequencies. Each cell is given an ID that is unique under the BSC controlling the tower. This is illustrated in Figure 4.2, where each color represents a different channel, and the cell ID is a four digit hexadecimal number. For simplicity, whenever I consider a radio tower in the rest of this thesis I will just write BTS, but it could just as well be a Node B.

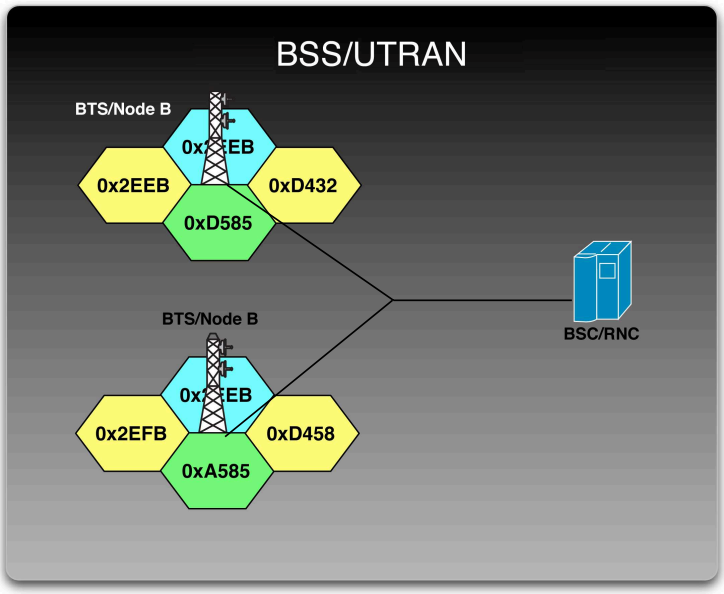


Figure 4.2: A simplified illustration of a BTS/Node B. The cell colors indicate frequencies and the texts are sample cell IDs.

Base Station Controller (BSC) / Radio Network Controller (RNC)

The BSC, or RNC, is a central unit connected to one or more radio towers and is the governing part of the radio networks in both standards. The main responsibilities are radio channel allocation and management of handovers between cells that are under their management. In addition, these units are responsible for merging multiple connections together and multiplexing them onto the channel to the Mobile Switching Center (MSC), as illustrated in Figure 4.1.

4.1.3 Core Network

The core network of a cellphone system is responsible for moving traffic from the switched telephone network onto the wireless network, and vice versa. In order to carry out these tasks, it needs to do a lot of bookkeeping, and every network maintains several registers and centers for bookkeeping purposes. I will now introduce two such registers used by the MSC, since they will turn out to be important to the privacy evaluation.

Home Location Register (HLR)

Every cellphone operator maintains a Home Location Register (HLR) containing detailed information about every subscriber. The database contains much of the same information as the customers' (U) SIM-cards. The most important use of this information is as part in the authentication process of subscribers. The HLR also stores the subscribers current position in the network. Every time a cellphone transfers from one cell to another, a location update message is sent to the subscribers HLR. Information about where the subscriber is located is used for routing calls and messages. When a call to a cellphone is placed, it is forwarded to the operator of that number. The operator then forwards the call to the HLR the call is addressed to, the HLR checks to see where the phone is located in the network, and forwards the call onto the correct BSC/RNC, which in turn connects it to the correct tower and the correct cell.

Visitor Location Register (VLR)

In the same way every MSC has a HLR, it also maintains a similar database of cellphones that have roamed into its area from other service providers. The Visitor Location Register (VLR) maintains information about the HLRs of all visitors, and use their HLRs to get authentication data, and for billing. As with the HLR, the VLR database is important when routing traffic to and from the UE.

4.2 Geolocalisation of Mobile Stations

This section gives the reader a general introduction to different methods available to determine the geographic position of a MS. It also includes a brief description of a Midlet I wrote to examplify geolocalisation from the MS using information from the cellphone network.

Simple Cell ID

The simplest way of approximating the position of a MS from the network is to determine which BTS it is currently connected to. Combining this with information about the geographical placement of the BTS, gives a general idea of which area the phone is currently located in. As earlier described, each BTS usually serves multiple cells. In most configurations directional antennas are used, so that each cell cover a directional cone, and not 360 degrees, typically a configuration utilizes three or six sectors. If one has

Cell type and coverage area.		
Cell type	Antenna location	Cell Radius (km)
Large macrocell	Above rooftop level	3-30
Small macrocell	Above rooftop level	1-3
Microcell	Below or about rooftop level	0.1-1
Picocell	Below rooftop level	0.01-1
Nanocell	Below rooftop level	0.001-0.0.1

Table 4.1: List over different cell types and their approximate area of coverage. The table was found in [43].

information about which cell the MS is connected to, where the radio tower is located, and the direction of the antenna, one can further determine an approximate position of the MS. A major problem with localization based on cell ID alone, is that the area covered by a cell can range from a small indoor area to an area with a 35 km radius. Table 4.2 lists different cell types and their typical size.

The cell ID serving the cellphone is the basis used by Google Mobile Maps to pinpoint the location of users who does not have a GPS chip in their phone. As a part of my thesis, I wrote a simple J2ME Midlet for my cellphone that fetched the information about where my phone was getting its service from. The Midlet fetches the Mobile Country Code (MCC) indicating the country I am in, the Mobile Network Code (MNC) giving information about the network operator, Location Are Code (LAC) indicating the area you are in, and the cell ID of the current cell. The sample Midlet is included in Appendix A, while a screen shot of the application is seen in Figure 4.3.

The numbers gathered from the phone is worthless without some more information about where the tower serving the specified cell is located. When I started working on my Midlet in the autumn of 2007, I could not find any such information, as the exact position of radio towers and which cells they are serving was considered secret by the Norwegian telephone network operators. But as Google released Mobile Maps, they obviously had access to the information. The API they used for determining the approximate position of a phone was soon reverse engineered, and released on the Internet. Using a simple php script, also included in Appendix A, I was able to use the Midlet and fetch an approximate latitude and longitude for the phone.

Timing Advance and Cell ID

Timing Advance (TA) is a GSM and UMTS network variable that is used to measure the time a signal uses from a BTS, to a cellphone and back again.

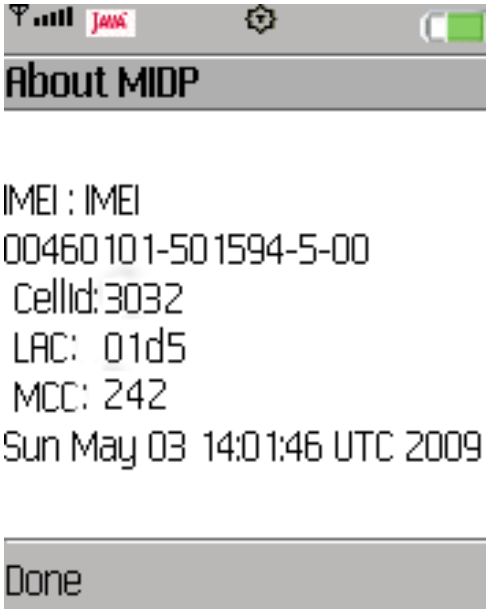


Figure 4.3: A screenshot of the Midlet fetching network information from a mobile phone.

In theory, TA could be used to calculate the distance between the BTS and a given cellphone. The problem with TA in GSM networks, however, is that the time is represented as a value between 0 and 63, with one step of this value representing roughly 3.69 microseconds. In this time a radio signal travels 1100 meters. Since TA measures round time we have to divide 1100 by two, giving TA a resolution of 550 meters. So assuming TA starts at 1, a TA value of 1 places the cellphone somewhere between 0 and 550 meters from the radio tower. A value of 2 places it between 550 and 1100 meters from the tower etc. Not very accurate when trying to determine where in the city you are currently located. UMTS employs a different radio technology, and has a better implementation of TA, resulting in a theoretical resolution of 35 meters, but according to [44] the resolution is much worse in practise.

Triangulation techniques

Various methods may be used to triangulate an object that emits a signal by observing the received signal from different places.

One such method is Time Difference of Arrival (TDOA). When using TDOA, the time of arrival of a specific signal is measured at different BTSs, the difference in this time between the base stations is used to determine the position of a MS. By looking at TDOA, the distance from one base to the MS

can be expressed as an hyperbolic. When several hyperbolas are calculated the intersection between them will reveal the approximate position of the MS.

Another triangulation technique is Angel of Arrival (AOA), a method where special antennas placed at the BTS are used to determine the angle of an incoming signal. By calculating multiple such angles and knowing the antennas' positions, the MS is located by "drawing" a line from each antenna in the calculated angle and looking at the intersection of these lines.

In addition, one could also imagine triangulation done with TA from three or more base stations. Using the TA value we get a radius from each BTS, and by drawing circles with these radiuses the MS is probably located somewhere inside the intersection of these circles.

But neither of these methods are applicable without adding expensive equipment to the network, so they are not much used in practice.

Global Positioning System

GPS is a positioning system based on satellites orbiting the earth, emitting radio signals. GPS enabled devices can receive this signal from satellites in sight and use it do calculate its position. This calculation is done by looking at the intersection of spheres.

As stated several high end phones also include a GPS chip. One major drawback of GPS however, is that it does not work well in urban areas, because it requires a line of sight to the satellites. In addition GPS can only be used from the handset, unless the phones come with functionality that allows the network to query them for location information.

Wireless Thumbprint

Many phones on the market today come with a WiFi chip, enabling them to use local WiFi hotspots for internet connections. Each wireless network consists of one or more wireless base stations, and each such base station has a unique address that it broadcasts over the air. The range of such networks are typically less than a 100 meters, thus knowing the position of the base station narrows your position down to a relatively small area. An American company, named Skyhook Wireless, are trying to make business from the idea by creating a big database with the positions of WiFi base stations. When a user then wants to determine his position he does a site survey, discovering nearby wireless networks, combines this with cell ID, and/or GPS data and sends it off to Skyhook. Skyhook Wireless then does a

lookup in their database and determines your position. For more information about this method see [45].

4.3 Privacy Management

Having given an introduction to the most important parts of the GSM/UMTS systems, and an introduction to geolocalisation, I will now turn to the framework introduced earlier in the thesis. The rest of this Chapter contains a sample privacy evaluation of GSM and UMTS with respect to traffic data and location information.

4.3.1 Collection

Any information system should only collect the minimum amount of personal information that it needs to fulfil its purpose.

The voice and data traffic has to make its way through the networks of the cellphone system. Although the traffic is encrypted when transmitted over the air, it travels in the clear through the wired networks. It should also be noted that the ciphers used to protect GSM have several proven weaknesses [46], and that commercially available equipment is capable of breaking it in real time [47].

SMS and MMS traffic are handled differently. Whenever the network receives a message that is to be delivered, it checks the HLR or the VLR, depending on where it finds the device. The information from these registers is used to determine the address of the Short Message Service Center, or the equivalent for MMS, associated with the recipient, and the system forwards the message there. The service center is then responsible for delivering the messages at a later time when the user is available.

Whenever an action that the user has to pay for is performed, whether it is sending a message, placing a call, or generating data traffic, a Call Detail Record (CDR) is generated. One thing that might not be as obvious at first, is that such records are also generated when a device receives traffic. So if two people are talking to each other using cellphones, a CDR will be recorded and stored for both of them, not just the one that placed the call. Each CDR contains at least the following information:

- The number placing the call.
- The number of the recipient.
- Date and time when the call was placed.

- Duration of the call.

- Type of call (Voice, Data, SMS etc.).

- The position of the parties (for those of the parties using a cellphone).

When discussing the phone network, it should also be noted that there are several strong identifiers in the system. A phone number is, necessarily, a unique identifier. In Norway, and most other countries, the registration process when buying a subscription aims to achieve a high grade of certainty that the person is who he claims to be. When buying a subscription one has to provide an officially approved ID, and the NBN is used to tie the phone number to a person. Another strong identifier is the International Mobile Equipment ID (IMEI), the IMEI is a unique 14 or 16 digit number, plus an additional checksum digit, that identifies each handset ever produced. Amongst other things, the IMEI number is used to block devices from the network in the case of a theft, and this number is broadcasted on the network by the phone. The third unique identifier used is the International Mobile Subscriber ID, which resides on the (U)SIM card which is used as an international identifier for GSM and UMTS subscriptions. So there are two unique identifiers that are tied to a subscription, and one that is tied to the mobile station.

To sum everything up, the cellphone networks contain information about you as a subscriber, they carry your voice and data communication unencrypted through their inner networks, SMS and MMS messages are at some point available in their respective message centers. In addition, they create and store CDRs every time you use your phone. There are also quite a few strong identifiers in the system, all of which are strongly connected with the subscriber through the initial registration process or to a specific phone (MS). All of these data points are created and collected in order to make the system function, and to enable billing of customers. The tight coupling between subscribers and phone numbers are in place due to regulations from the government.

As outlined in italics in the beginning of the section, a system should not store more information than it needs to carry out the tasks it was built for. The information collected by cellphone systems is in large needed in order to do billing and network planning. However, I have not found a very good explanation as to why the geographical position of the user is stored in the CDRs. As a result, I would rate Collection as medium.

4.3.2 Retention

Personal information should be retained for the shortest possible period.

The phone companies need to store CDRs in order to do billing, both towards their own customers, but also in order to charge other companies for roaming visitors. In addition, traffic data may also be used for network-planning. So the phone companies obviously need to store CDRs for some amount of time in order to get paid. Norwegian law mandates that companies are to delete, or anonymize, such traffic data as soon as they are not needed for communication or billing purposes, unless otherwise specifically instructed by law. In any case such information should be deleted within three months if the subscriber is paying monthly, and after five months if the subscriber pays every quarter, unless there are special circumstances such as a dispute or a police investigation, according to the Norwegian Data Inspectorate [48].

For pre-paid subscriptions there is no need to store data for billing purposes, data only needs to be stored long enough for the customer to place a complaint. There have been some speculations that the transition we are seeing towards different prepaid, or "call as much as you want," subscriptions is a part of the reason for the introduction of the EU data retention directive. If traffic data is not needed anymore, then phone companies may stop storing it, thus ruining an important source of information for law and intelligence agencies.

For retention I suggest a rating where the criterion for high is as outlined in italics at the beginning of this section. In order to get a medium rating, retention times should be slightly longer than strictly needed, but still data should be deleted on a regular basis. Low should be given in cases where information is retained for a longer, or even, indefinite amount of time.

In theory, and according to legislation in Norway, the retention time for meta data about phone activity should be rated as high. The data are only stored for billing purposes and are to be deleted after three to five months, depending on the individual payment plan. In practice however the image seems to be somewhat different. As a part of my research, I contacted my provider, Chess, to request a printout of my traffic information. During the correspondence I was informed that such data was available for 150 days back in time, even though I have a monthly payment plan. Another fellow student of mine, discovered that he had access to specified bills dating back to 2007 on his cancelled landline account with the Norwegian provider NextGenTel. When he contacted customer services inquiring why this information had not been deleted, he got a reply stating that they had no obligation to delete the information. Puzzled by the reply he contacted the Norwegian Data Inspec-

torate. They sent us a copy of the concession NextGenTel had for processing this information, which clearly stated that they were obligated to delete this information. The Data Inspectorate also contacted NextGenTel and instructed the company to implement routines for deleting such information, and to keep the Inspectorate informed about their process.

So even though I have rated retention as high based on laws and publicly available information, if I were to do a rating of my provider Chess, they would fall down to medium, while NextGenTel who obviously do not have routines for deletion at all would end up with a rating of low. If the data retention directive is implemented in Norway, I would have rated any system storing traffic data for more than six months as medium or low, depending on the storage time; simply because of the dramatic increased amount of information that would be residing in the system, and the long storage time. In the worst case, granted that you use your cellphone a few times a day, a complete map of your movements for the last two years would be available to anyone with access to your records.

4.3.3 Secondary Use

Collected personal information should only be used for the specific purpose it was originally collected.

According to the Norwegian Law, any use of CDRs that goes beyond billing requires an active consent from the subscriber. So here there are good juridical restrictions on the use of traffic data protecting subscribers from function creep.

If I were to define the high, medium and low criterions, I would phrase high as the information is strongly protected against secondary use. Medium would be for systems where protection against secondary use are present, but not working in an optimal way. While low would be for systems that either facilitate secondary use, or do not have good policies or agreements with with customers concerning their personal information.

From these definitions, the protection against secondary use in the Norwegian telephone system should be rated as high.

4.3.4 Distribution

The personal information collected by any system should not be made available to third parties without prior consent from the data subject.

As with secondary use, external sharing of information also falls under the laws requiring the consent of those registered in the system, but there are some exceptions for emergency services and police investigations. Every

call that is made to a Norwegian emergency department (police, ambulance, or fire department) from a cellphone is automatically positioned and the approximate location of the caller is displayed on a map in front of the operator. Revealing the position of a caller in distress is actually one of the main reasons why the functionality for locating a MS geographically was developed, and extended the way it has. Before cellphones, every landline was tied to a specific address, so the location of the caller was simply a matter of looking in the phone book. But with the introduction of the cellphone, emergency responders were dependent on the caller to know their location. As a result governments required the telephone companies to enable geographic positioning of cellphones.

The police is also granted access to traffic data if they are conducting an investigation. The only requirement is that they are conducting an investigation and forward a request to the Norwegian Post and Telecoumications Authority describing their needs, no court ruling is needed [49].

Having built the geolocalisation functionality into the cellphone network, providers also want to profit from it. By offering other companies access to the location of mobile devices, providers are able to make money on such services. Of course, this access is regulated by agreements stating what is allowed and what is needed in order to determine the position of a MS. While working on the thesis I contacted one company that I knew had such an agreement and tried to get more information about it. They said that in order to trace someone the user had to consent to the tracing before it is carried out. Further, I asked whether there were any technical functionality that enforced this rule or if it was just a contractual thing, they did not reply. If this is in fact the case it opens for an insider threat, where anyone with access to the tracking functionality are able to track someone without prior consent.

The information about your position can of course also be disclosed by software running on your cellphone, as explained and exemplified at the beginning of this Chapter. Applications such as Google Mobile Maps send the information about where in the network you are connected when pinpointing your location. Of course the network providers are not to blame for it, but it should be noted that with the right software installed on the MS it is possible to track the MS without involving the phone company. American law enforcement used malicious software to turn mobile devices into spying devices according to a news article on CNET [50]. There is even software commercially available for turning a cellphone into a complete surveillance tool, allowing everything from remote listening, controlling the phone, and determining its position. One example of such a software is FlexiSpy [51].

In order to rate as high, it should not be possible for anyone to get hold

of personal information without prior consent. We have seen that Norwegian legislation in large aligns with this requirement, except in the case of the police, who are free to collect traffic information related to an investigation. Medium should be used for systems where personal information is distributed to other parties on a regular basis, but the distribution is well documented and based on real needs. A rating of low should be given to systems where prevention, or at least documentation, of information distribution is not present.

Following these guidelines, the personal information maintained in the cellphone system in Norway gets a rating of medium, due to the fact that the Norwegian Police can access to traffic data without a court ruling, and the possible leakage of users positions through insider threats in companies with B2B agreements about access to location information.

4.3.5 Distortion

Operators of an information system should do their best to assure the integrity of the system.

The personal information in the cellphone system is largely used for billing of customers and other telephone companies. As a result, the companies themselves have good incentives to make sure that information in the system cannot be manipulated in any way. It is worth noting that messages sent in the phone system does not provide any checking for integrity, and it is easy to spoof the sender number if one has access to an SMS gateway. Likewise it is possible to spoof the caller ID, a hack that was used for listening to Paris Hilton's voicemail and to steal her phonebook a few years ago. Doing so was possible because the voicemail did not have any password, it just checked the callers number, if there was a voicemail account for that number access was granted.

In addition, if we agree that traffic data is indeed personal information, paragraph thirteen in the Norwegian law regulating the processing of personal information talks about information security. It states that anyone who processes personal information are to ensure that the system offers an adequate level of security with regards to confidentiality, integrity, and availability through a planed and systematic effort. Further, it states that the effort should be documented, and the documentation should be made available to the Data Inspectorate. Although the law does not further specify what is considered as an adequate level of security.

In order for a system to get rated as high with respect to distortion, the system has to incorporate security in its design, and the safety measures in place to protect personal information must be well documented and publicly

available. Systems should be granted a rating of medium when the efforts to ensure integrity are not adequately implemented, or not well documented. Low should be given to systems where efforts to ensure the integrity of personal information are not present, or left undocumented.

According to these criterions I would give the cellphone system a high rating when speaking of call-details and position only. If the integrity of other types of personal information, such as messages and voicemail, were to be included the rating would probably be lower.

4.3.6 Correction

Any individual about whom the system stores personal information should be able to access and correct data concerning self.

Since the information is collected automatically in the phone system and used as the basis for billing, there should be routines in place that allow customers to complain if they think that their bill is somehow wrong. In order to place a call using someone's subscription in a GSM or a UMTS network one essentially needs two numbers that reside on the (U)SIM card. However, a smart card is built to be a tamper resistant device that goes a long way to protect the confidentiality of the information that resides on them. As to my knowledge there have not yet been mounted any practical attacks on the current version of these cards that allow an attacker to extract this information using methods that does not destroy the SIM card.

But no matter how good the technical protection built into the system, customers should still be able to access the information concerning themselves. If the information is erroneous, routines for placing a complaint and correcting it should be in place. Norwegian telephone operators will give you access to the information stored in the system about your subscription if they are addressed in writing by the subscriber. As they are obliged to do so by the law regulating the use of personal information, as described in the "Correction" Paragraph in Section 3.2.2.

However after addressing a formal letter to my provider at the time, which was Chess, asking for information about the security measures, and a list of traffic data, they replied with a list of incoming traffic and failed to account for the security measures and to describe what types of personal information they were processing. Drawing any conclusions based on just one incident is impossible, but at least my provider seems to have some room for improvement.

In order for a system to get rated as high there has to be formalized routines for how to access ones personal information, and how to deliver a complaint, or ask for information to be corrected. While I would rate a system

where you are still allowed to do these things, but not in a simple and formalized way, as medium. For systems where access to ones personal information is not granted, the control should be given a rating of low. In Norway any systems rated below high are, as I see it, in violation with sanctioned laws, and should implement processes to accommodate the requirements made by the law.

4.3.7 Notification

In the case of a mishap the users whose personal information was leaked and, perhaps, misused should be informed of the incident.

When personal information is lost or stolen, then those affected should be informed so that they may take precautionary steps to protect themselves. I have tried to find information about how notification would be given on the public web pages of telephone providers in Norway. But I have not found any information about how they would handle such an event. So I have chosen to label this control as missing, since I was not able to find any notification information, neither did I find anything in the press where phone companies have warned about leakages. Whether the lack of such stories imply that they do no exist or that they are not publicly disclosed would only be speculation.

In America, most of the states have sanctioned laws requiring the public release of information about data breaches. In some states the law requires that those affected by the breach are informed in writing, stating that a public disclosure of the breach is not enough. Norway on the other hand seems so largely ignore data breaches, the law however clearly states that personal information should not be shared with anyone. Except in the case where information is collected with the purpose of sharing it, for instance credit ratings.

4.3.8 Summary

In this Chapter I have done an external review on the privacy of traffic information in Norwegian cellphone networks. Overall, the information is granted good protection by the Norwegian law, and it seems that the phone companies are very much in compliance with regulations when it comes to meta data about communication and location in general. Figure 5 shows us that in theory the overall privacy of traffic data is well taken care of. However, as I have explained when evaluating the different controls, several of the companies providing services in Norway would get a lower rating if reviewed individually; indicating that conducting such a review could increase the privacy these companies offer their customers. In the case of NextGeTel,

a review would have detected that they are currently operating in clear violation of Norwegian law by not deleting the specified bills that they issue to their customers.

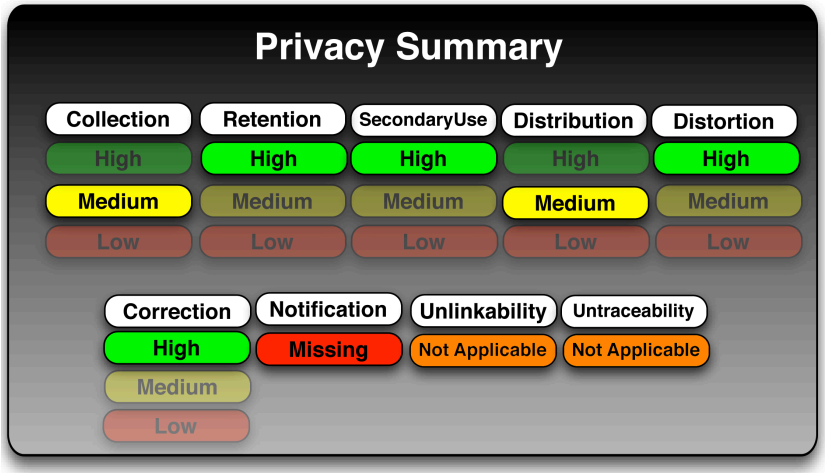


Figure 4.4: A simple illustration summing up the review.

Chapter 5

Summary and Conclusions

5.1 Summary

This thesis started by introducing privacy as a concept and argued why privacy of personal information is important. We saw that certain types of personal information was considered sensitive, and granted special protection by the law. In addition, we have seen that the Norwegian law places restrictions on the use and storage of personal information. So regardless of whether an organization thinks privacy is important or not, it has to provide adequate privacy whenever building or maintaining information systems where personal information is present.

Chapter 3 gave an introduction to risk management. Risk management was then used as inspiration for a method that could be used to evaluate the overall privacy in an information system, based on nine different controls. After the introduction of this method, an example privacy review was carried out on the cellphone system with focus on traffic and location data. The review revealed that the privacy of such data is in theory well protected, but there is still some room for improvement as two controls were labelled medium and one as missing. Additionally it was discovered that several telephone providers did not have good routines for deletion of personal data.

A sample Java 2 Micro Edition Midlet was developed to demonstrate geolocalisation of a mobile phone. The Midlet includes functionality for approximating the devices position using information gathered from the network along with services from Google, as well as the possibility to push information about its location to a web server. Thus, enabling a third party to follow the movements of the device.

5.2 Subjectivity in Results

It should be noted that the sample evaluation of the cellphone network in Chapter 4 is based on information freely available and was carried out by the author alone. If more information had been available, or if others had joined in on the evaluation, the results may have been different, but it is the authors view that the conclusions drawn in the evaluation are founded on the material available and the sources referenced.

When evaluating privacy, as with risk assessments, the best results will emerge when carried out by a team of individuals with different experiences and skills.

5.3 Criticism

The framework suggested for evaluating privacy in this thesis is largely based on work by me and my advisor, and additionally inspired by the American Privacy Impact Assessments mentioned at the end of Chapter 2. The presented framework should be viewed as a first attempt to better measure privacy in information systems. While the reader may question the usefulness of yet another framework, it is my view that privacy will be an increasingly important topic in the years to come, and little work is available on how to actually measure privacy. Hopefully, this will change and we will see a rise both in techniques developed for preserving privacy in different settings and places, as well as more research on how to preserve the overall privacy of systems.

5.4 Conclusions

By using the proposed framework to evaluate parts of the cellphone network, we were able to identify areas where privacy could be improved. When looking closer at how different providers were storing their traffic data it was discovered that at least one of them was operating in clear violation of their concession, while the other one also seemed to violate their concession, because they stored data for more time than permitted to. So it seems that trying to manage privacy by carrying out a structured analyze of a system would be beneficial both for customers and companies.

When large companies as Chess/NetCom and NetGenTel are found to break the law by outsiders analyzing their systems it suggests the need for more privacy analyses of Norwegian companies. Hopefully, in the future we

will se even more pressure from the Data Inspectorate and others to enforce privacy laws and regulations. After all laws are useless without compliance.

5.5 Further Work

More work on the controls would probably be beneficial, since only a handful of individuals have helped in the process of establishing and refining these. Perhaps it could be possible to determine some general criteria for high, medium, and low, for use on all systems.

Additionally, applying the method to multiple large systems would probably be very useful in future refinements. It would also be interesting to do reviews of how specific telephone providers in Norway are handling traffic and location data. Research of how much information could be extracted from two years worth of such records would also be a valuable input to the debate about the data retention directive.

Bibliography

[1] "ISC Internet Domain Survey, Jan 2009," URL http://ftp.isc.org/www/survey/reports/2009/01/.

[2] "Prosjektdirektiv for e-valg 2011," Technical report, Komunal- og Regionaldepartementet, 2009.

[3] "Electronic voting – challenges and opportunities," Technical report, Norwegian Ministry of Local Government and Regional Development, 2006, URL http://www.regjeringen.no/upload/kilde/krd/red/2006/0009/ddd/pdfv/272294-elektroniskstemmegivning.pdf.

[4] L. D. Brandeis and S. D. Warren, "The Right to Privacy," *Harvard Law Review*, IV, December 1890.

[5] Personvernkommisjonen, "NOU 2009:1 Individ og integritet - Personvern i det digitale samfunnet," Technical report, Norwegian Ministry of Government Administration and Reform, January 2009.

[6] Privacy International, "Overview of Privacy," Technical report, Privacy International, 2007.

[7] B. Schneier, "The Eternal Value of Privacy," Comment on Wired.com, May 2006, URL http://www.wired.com/politics/security/commentary/securitymatters/2006/05/70886.

[8] G. Orwell, *1984*, Signet Book, May 1990.

[9] C. Doctorow, *Little Brother*, Tor Teen, 2008.

[10] Department of Homeland Security, "Fact Sheet: CAPPS II at a Glance," Press Release, 2004.

[11] J. Lettice, "Automated Profiling Tech is Crap, says Home Office," TheRegister.co.uk, June 2008.

[12] S. Gaycken, "Arguments Against Surveilance," 2007, URL http://events.ccc.de/camp/2007/Fahrplan/track/Society/2021.en.html.

[13] H. A. Landsberger, *Hawthorne revisited : Management and the worker: its critics, and developments in human relations in industry*, Cornell studies in industrial and labor relations, Ithaca, N.Y. : Cornell University, 1958.

[14] G. Apenes, "Har Personvernet Noen Fremtid? (Is There Any Future For Privacy?)," September 2008, URL http://digi.no/php/art.php?id=788020.

[15] J. King, D. K. Mulligan, and S. Raphael, "An evaluation of The Effectiveness of San Francisco's Community Safety Cameras," Technical report, he Center for Information Technology Research in the Interest of Society, December 2008.

[16] M. Gill and A. Spriggs, "Assessing the Impact of CCTV," Technical report, Home Office Research, Development and Statistics Directorate, 2005.

[17] Personvernnemda, "Klage på vedtak om begrensninger i adgangen til kameraovervåking av bussenes publikumsområder." August 2006.

[18] I.-A. Ravlum, *Setter vår lit til Storebror ... og alle småbrødre med?*, Transportøkonomisk Institutt, 2005.

[19] "The European Data Protection Directive," 2005.

[20] "Lov om behandling av personopplysninger (personopplysningsloven)." Norwegial Legislation, April 2004.

[21] Symantec, "Symantec Report on the Underground Economy," Technical report, Symantec Corp, Novermber 2008.

[22] S. T. Kent and L. I. Millett, *Who Goes There?: Authentication Through the Lens of Privacy*, National Research Council Washington, 2003.

[23] A. Klingsheim, "Identity Theft: Much too Easy?" in *Financial Cryptography and Data Security, 12th International Conference, FC 2008,*

Cozumel, Mexico, January 28-31, 2008, Revised Selected Papers, volume 5143 of *Lecture Notes in Computer Science*, pp. 42–42, Springer, 2008.

[24] E. S. Selmer, "Personnummerering i norge: Litt anvendt tallteori og psykologi." *Nordisk Matematisk Tidsskrift*, 12:pp. 36–44, 1964.

[25] R. L. Mitchell, "What the Web knows about you," *Computerworld*, 2009.

[26] URL http://www.privacyrights.org/.

[27] URL http://www.dhs.gov.

[28] G. Stoneburner, A. Gougen, and A. Feringa, "Risk Management Guide for Information Technology Systems," *NIST Special Publication 800-30*, 2002.

[29] G. McGraw, *Software Security—Building Security In*, Addison-Wesly Publishing, 2006.

[30] M. Howard and S. Lipner, *The Security Development Lifecycle*, Microsoft Press, 2006.

[31] L.-H. Netland, *Assessing and Mitigating Risks in Computer Systems*, Ph.D. thesis, University of Bergen, 2008.

[32] T. Aven, *Risikostyring*, Universitetsforlaget, 2007.

[33] N. N. Taleb, *The Black Swan: The Impact of the Highly Improbable*, Random House, April 2007.

[34] K. J. Hole, "Toward Risk Assessment of Large-Impact Events," 2009, unpublished manuscript.

[35] P. Syverson, M. Reed, and D. Goldschlag, "Onion Routing Access Configurations," in *Proceedings of the DARPA Information Survivability Conference and Exposition (DISCEX 2000)*, volume 1, pp. 34–40, IEEE CS Press, 2000.

[36] A. Strand and L. Bertheussen, "Flexus Under Lupen," *NRK - Østlandet*, 2009.

[37] M. Babaro and T. Zeller, "A Face Is Exposed for AOL Searcher No. 4417749," *New York Times*, 2006.

[38] R. Brownsword, "Knowing Me, Knowing You - Profiling, Privacy and the Public Interest," in *Profiling the European Citizen*, pp. 345–363, Springer Netherlands, 2008, URL http://www.springerlink.com/content/v74275w74262630g/.

[39] J. Smith, "Report on the Operation in 2007 of the Terrorism Act 200 and of Part 1 of the Terrorism Act 2006," The Stationary Office, 2008.

[40] B. C. Peake, "Red Teaming: The Art of Ethical Hacking," Whitepaper, 2003.

[41] 1998, URL http://catless.ncl.ac.uk/Risks/19.97.html#subj2.

[42] 1998, URL http://catless.ncl.ac.uk/Risks/20.02.html#subj4.

[43] S. Y. Willassen, *A method for implementing Mobile Station Location in GSM*, Master's thesis, Norwegian University of Science and Technology, 1998.

[44] B. Denby, "Geolocalisation in Cellular Telephone Networks," in *NATO Advanced Study Institute on Mining Massive Data Sets for Security*, 2007.

[45] Skyhook Wireless, "Techrapport from Skyhook," Technical report, Skyhook Wireless, 2006.

[46] A. Biryukov, A. Shamir, and D. Wagner, "Real Time Cryptanalysis of A5/1 on a PC," in *In FSE: Fast Software Encryption*, pp. 1–18, Springer-Verlag, 2000.

[47] T. Güneysu, T. Kasper, M. Novotný, C. Paar, and A. Rupp, "Cryptanalysis with COPACOBANA," *IEEE Trans. Comput.*, 57(11):pp. 1498–1513, 2008.

[48] The Norwegian Data Inspectorate, "Concession for processing of personal information—Processing of information about subscribers use of communication services." Standard Contract for Service Providers in Norway.

[49] The Norwegian Post and Telecommunications Authority, "Begjæringer om fritak fra tilbyders lovpålagte taushetsplikt," Technical report, The Norwegian Post and Telecommunications Authority, 2007.

[50] D. McCullagh and A. Broache, "aps cell phone mic as eavesdropping tool," *CNET News.com*, 2006.

[51] URL http://www.flexispy.com.

[52] URL www.opencellid.org.

[53] "Developer Guidelines—Java Platform, Micro Edition, CLDC - MIDP 2 for Sony Ericsson feature and entry level phones," Technical report, Sony Ericsson, 2009.

Appendix A

Sample Cell ID Midlet

At the beginning of my work on this thesis I came across an J2ME Application Programming Interface (API) from Sony Ericsson that allowed programmers to fetch information about where in the network the phone was connected. Using this information, it would be possible to determine the approximate position of the cellphone, granted one had access to a database of BTS and Cell ID locations. So to demonstrate how this could be done, I wrote a sample Midlet fetching this data from the phone and pushing it to a central server.

At the time, I did not have access to any databases where I could look up the position of the specific cell, and I discovered that this information was allegedly considered secret by Norwegian providers. Just a few months later, Google released their Mobile Maps application, including a feature to locate phones without GPS chips. I quickly realized that they were using positioning based in cell ID. In the following sections, I will outline the API provided by Sony Ericsson, how I used Google to look up the position of cells, and finally the source code for a simple MIDlet is included.

As a side note, the open source project "OpenCellID" aims at creating a complete database of mobile cells and their geographical placement [52].

A.1 Sony Ericsson API calls

Below is a short list of the most interesting properties that the Sony Ericsson Java Platform allows users to fetch using the System.getProperty(" "); call. All the commands work with version 7.3 of Sony Ericsson Java platform, except for the last two who are only supported in version 8 and later. Information about these, and other, calls can be found in the Sony Ericsson

Developer Guidelines for the Java Platform, Micro Edition, CLDC - MIDP2 [53].

com.sony Ericsson.net.mcc Used to fetch the home mobile country code from the phone. For instance in Norway this would be 242.

com.sony Ericsson.net.mnc Used to fetch the home mobile network code. In Norway 01 is allocated to Telenor, 02 is NetCom. The code is usually two or three digits long.

com.sony Ericsson.net.cmcc Used to fetch the country code of the network you are currently connected to.

com.sony Ericsson.net.cmnc Fetches the current mobile network core.

com.sony Ericsson.net.isonhomeplmn Returns true or false.

com.sony Ericsson.net.rat Used to determine the Radio Access Technology (RAT) currently used (GSM or WCDMA).

com.sony Ericsson.net.cell ID Returns the identity of the cell your phone is currently connected to in hex format. The value is four digits for GSM networks and eight for WCDMA.

com.sony Ericsson.net.lac Returns the local area code where your phone is connecting to the network.

com.sony Ericsson.net.serviceprovider Returns the name of the operator or service provider.

com.sony Ericsson.net.networkname Returns the name of the network the phone is connected to.

With these calls I was able to fetch all the information that is needed to geographically locate a device using the simple Cell Id technique described earlier in this thesis.

A.2 Google Mobile Maps API hack

As stated at he beginning of the Appendix, a database with necessary information about cells and their geographical placement was not available when I wrote the first version of this Midlet. But when Google released their Mobile Maps, I found a discussion on a Google forum, where the users were attempting to reverse engineer the API used to turn the information about

where the phone was connected to the phone network into geographical coordinates. After a short period of time, the API call needed to do this was reverse engineered, and available on the net.

So, using the following php code on a web server i controlled, I was able to input the needed data and having Google return the latitude and longitude.

```
$init_pos = strlen($data);
$data[$init_pos - 38]= pack("H*",substr($mnc,0,2));
$data[$init_pos - 37]= pack("H*",substr($mnc,2,2));
$data[$init_pos - 36]= pack("H*",substr($mnc,4,2));
$data[$init_pos - 35]= pack("H*",substr($mnc,6,2));
$data[$init_pos - 34]= pack("H*",substr($mcc,0,2));
$data[$init_pos - 33]= pack("H*",substr($mcc,2,2));
$data[$init_pos - 32]= pack("H*",substr($mcc,4,2));
$data[$init_pos - 31]= pack("H*",substr($mcc,6,2));
$data[$init_pos - 24]= pack("H*",substr($cid,0,2));
$data[$init_pos - 23]= pack("H*",substr($cid,2,2));
$data[$init_pos - 22]= pack("H*",substr($cid,4,2));
$data[$init_pos - 21]= pack("H*",substr($cid,6,2));
$data[$init_pos - 20]= pack("H*",substr($lac,0,2));
$data[$init_pos - 19]= pack("H*",substr($lac,2,2));
$data[$init_pos - 18]= pack("H*",substr($lac,4,2));
$data[$init_pos - 17]= pack("H*",substr($lac,6,2));
$data[$init_pos - 16]= pack("H*",substr($mnc,0,2));
$data[$init_pos - 15]= pack("H*",substr($mnc,2,2));
$data[$init_pos - 14]= pack("H*",substr($mnc,4,2));
$data[$init_pos - 13]= pack("H*",substr($mnc,6,2));
$data[$init_pos - 12]= pack("H*",substr($mcc,0,2));
$data[$init_pos - 11]= pack("H*",substr($mcc,2,2));
$data[$init_pos - 10]= pack("H*",substr($mcc,4,2));
$data[$init_pos - 9]= pack("H*",substr($mcc,6,2));

if ((hexdec($cid) > 0xffff) && ($_REQUEST["mcc"] !=
    "") &&
   ($_REQUEST["mnc"] != "")) {
   $data[$init_pos - 27] = chr(5);
} else {
   $data[$init_pos - 24]= chr(0);
```

```
  $data[$init_pos - 23]= chr(0);
}

$context = array (
        'http' => array (
            'method' => 'POST',
            'header'=>
              "Content-type: application/binary\r\n"
                    . "Content-Length: " . strlen(
                       $data) . "\r\n",
            'content' => $data
            )
        );

$xcontext = stream_context_create($context);
$str=file_get_contents
  ("http://www.google.com/glm/mmap",FALSE,$xcontext)
     ;
$lat = ((ord($str[7]) << 24) | (ord($str[8])
   << 16) | (ord($str[9]) << 8) | (ord($str[10])))
    / 1000000;
$lon = ((ord($str[11]) << 24) | (ord($str[12])
   << 16) | (ord($str[13]) << 8) | (ord($str[14])))
     / 1000000;
```

A.3 The Midlet

The Midlet I wrote has two basic functionalities. If the user selects "Start pushing" the cellphone will collect information about where it is connected in the network and push this onto a web server along with the IMEI of the cellphone. At this web server the information is stored in a database and can be used in different ways. To illustrate one possible use, a simple interface displaying the last recorded positions of a phone was implemented, a screenshot of this is seen in Figure B.1.

When the user selects "System Info" a screen is displayed for a few seconds where he is presented with a summary of the network information. If the user wants to, the phone may also fetch the latitude and longitude from the internet using the GMM API. A sample screen shot of this is seen in Figure

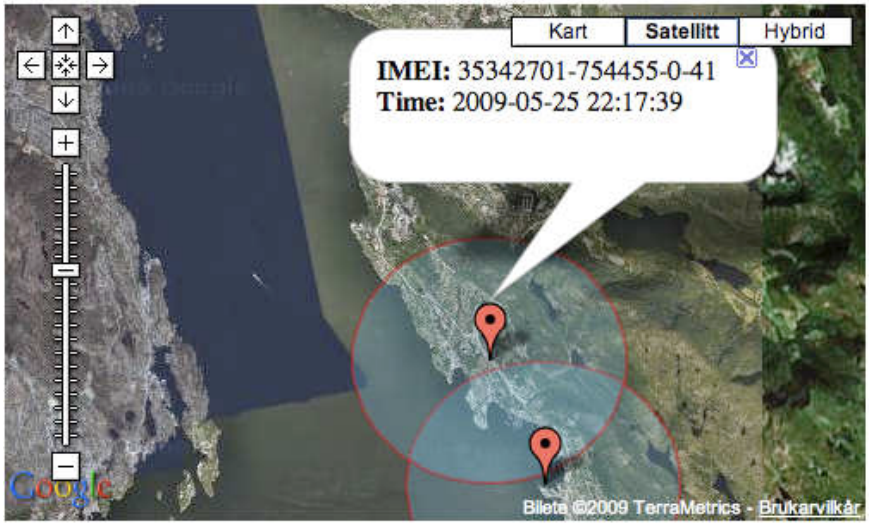


Figure A.1: A screenshot from the simple web page written by the author, as one possible use of the data pushed by the cellphones.

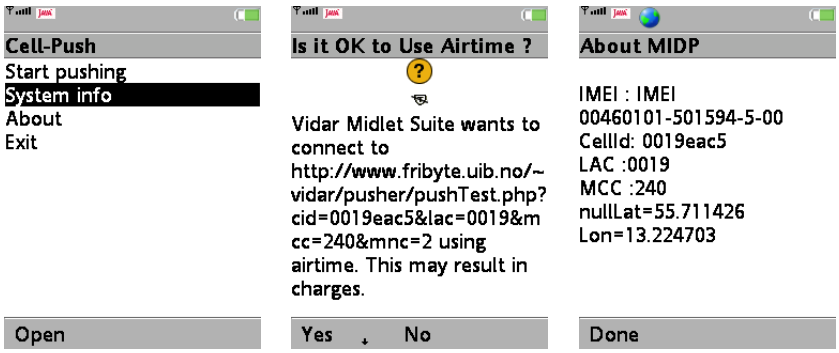
A.3

A.3.1 GeoLocalisationMidlet

This is the Class containing the main menu, and contains the functionality needed to set up the different screens.

```
import javax.microedition.lcdui.Alert;
import javax.microedition.lcdui.Command;
import javax.microedition.lcdui.CommandListener;
import javax.microedition.lcdui.Display;
import javax.microedition.lcdui.Displayable;
import javax.microedition.lcdui.List;
import javax.microedition.midlet.MIDlet;
import javax.microedition.midlet.
    MIDletStateChangeException;

/**
 * The MIDlet class of an application that fetches
     information about network
 * connectivity and uses this to determine the
     geographic position of the device.
```



(a) The main menu of the Midlet.

(b) The applications asks permission to connect to the Internet.

(c) A screen listing network information as well at latitude and longitude.

Figure A.2: Here we see the the different steps needed in order to view the network information in the Midlet.

```
 * It also supports pushing of the network
    information to a remote server,
 * enabling tracking of the cellphones whereabouts.
 *
 * @author Vidar Drageide
 *
 */
public class GeolocalisationMidlet extends MIDlet
   implements CommandListener {


  public List menuList ;        // Will contain the
     main menu
  private Command selectCommand ; // Command used to
     select items

  private Display disp ;        // Hook to the device
     display


  /**
   * Constructor
```

```
 */
public GeolocalisationMidlet() {

  // Fetch the display of the mobile device
  disp = Display.getDisplay(this);

  // create the menu for the application
  String[] elements = {"Start pushing","System
     info", "About", "Exit" };
  menuList = new List("Cell-Push", List.IMPLICIT,
     elements, null );

  // Add commands
  selectCommand = new Command("Open", Command.ITEM
     , 1);
  menuList.setSelectCommand(selectCommand);
  menuList.setCommandListener(this);

  // Display the menu list on the screen
  disp.setCurrent(menuList);
}


/**
 * This method is called when the application is
    shut down.
 */
protected void destroyApp(boolean arg0) throws
   MIDletStateChangeException {
  notifyDestroyed();
}

/**
 * This method is used in the case of a blocking
    event from the cellphone
 * that needs to put the Midlet at pause.
 */
protected void pauseApp() {
```

```
}

/**
 * This method is called to start the application
    after receiving a
 * blocking call.
 */
protected void startApp() throws
   MIDletStateChangeException {

}

/**
 * This method is called whenever a command is
    issued to the midlet
 *
 */
public void commandAction(Command arg0,
   Displayable arg1) {

  if(arg0.equals(selectCommand)){
    // Get the index of the element selected in
       the menu
    int key =  menuList.getSelectedIndex() ;
    try{
      switch (key) {
      // 0 - User selected start push
      case 0: startPush() ;      break ;
      // 1 - User selected Show system info.
      case 1: systemInfo();      break ;
      // 2 - Users selects about. Invoke the "
         About"-screen
      case 2: startAbout();    break ;
      // 3 - User wants to exit the software
      case 3: destroyApp(true);    break ;

      default: System.out.println("Default");
      break;
      }
```

```
        }
        catch (Exception e) {
          System.exit(1); // Exit with error
        }
    }
}

/**
 * Method that is used to set up everything needed
     for pushing information
 * to the web-server.
 */
private void startPush() {
  System.err.println("Starting to push information
      to server");
  IdPush pusher ;
  Thread pushThread ;

  pusher = new IdPush(this, disp);

  try{
    pushThread = new Thread(pusher);
    pushThread.start() ;
  }
  catch (NullPointerException e) {
    System.out.println(e.getMessage());
  }
}

/**
 * This method is invoked when the user presses "
    System Info"
 * The method fetches all the information that is
    to be displayed and
 * fetches latitude and longitude from the
    Internet.
 * The screen goes away by itself 10 seconds after
     answering yes
 * or no to "grant Internet access" dialogue.
```

```
 */
private void systemInfo() {

  System.err.println("Starting SystemInfo Screen")
     ;
  Alert alert = new Alert("About MIDP");
  alert.setTimeout(Alert.FOREVER);
  String imei = System.getProperty("com.sony
     Ericsson.imei");
  String cell ID = System.getProperty("com.sony
     Ericsson.net.cell ID");
  String lac = System.getProperty("com.sony
     Ericsson.net.lac");
  String mcc = System.getProperty("com.sony
     Ericsson.net.cmcc");
  String mnc = System.getProperty("com.sony
     Ericsson.net.cmnc");

  /*
   * Fetch latitude and longitude using php-code
      on a server
   */
  GeoFetcher a = new GeoFetcher(this);
  a.setCID(cell ID);
  a.setLAC(lac);
  a.setMCC(mcc);
  a.setMNC("2");
  a.setAl(alert);
  a.setDisp(disp);

  /*
   * Whenever the MIDlet needs access to the
      network
   * the phone opens a dialog querying the user
      whether
   * to allow this or not.
   * To avoid a deadlock this process should be
      separated into
   * a own thread.
```

```
        */
        Thread sjekkeThread = new Thread(a);
        sjekkeThread.start();

        // While we are waiting for the position we
           display the info we have
        alert.setString("IMEI : " +imei + "\n" + "Cell
           ID: " + cell ID + "\n"
            + "LAC :" + lac + "\nMCC :" + mcc + "\n" + a
               .location);
        disp.setCurrent(alert);
    }



    /**
     * Method that invokes the "About"-screen.
     * This is just at really simple about screen.
     */
    private void startAbout() {
        // Create an alert
        Alert alert = new Alert("About MIDP");
        // Display it until the user clicks "Done"
        alert.setTimeout(Alert.FOREVER);
        // Fetch the display
        Display display= Display.getDisplay(this);
        // SEt up and display the alert
        alert.setString("CellIDPush \nWritten by: Vidar
           Drageide " +
            "\nwww.drageide.com " );
        display.setCurrent(alert);
    }
}
```

A.3.2 IdPush.java

This class implements a simple push functionality. It collects and remembers the latest cell ID's and the time where the phone was last connected. When-

ever the class detects that the phone has roamed into a new cell it does a get-request to a web server. This request sends information about the phone IMEI, current cell ID, and time to the server.

```
package com.drageide.geolocalisation.CellIDPusher;

(Imports removed)

public class IdPush implements Runnable{
  private int CELL_MEMORY = 10 ;

  // Parent MIDlet, used for fetching the main menu
     if need be
  private GeolocalisationMidlet parent ;
  // Hook to the display of the device
  private Display disp ;

  private Form page ;

  // Array for storing location data.
  private Position[] loc ;

  // String used for storing data ;
  private String currentCell ;
  private String imei ;

  // Exit and back commands.
  private final Command EXIT_CMD = new Command("Exit
     ", Command.EXIT, 2);
  private final Command BACK_CMD = new Command("Back
     ", Command.BACK, 1);


  /**
   * Constructor.
   * @param d reference to the parent midlet.
   * @param disp hook to the display of the device
   */
```

```
public IdPush(GeolocalisationMidlet d, Display
    disp){
  try{
    loc = new Position[CELL_MEMORY];
    this.parent = d ;
    this.disp = disp ;
    this.imei = System.getProperty("com.
       sonyericsson.imei");
    imei = imei.substring(5);
  }
  catch (Exception e) {
    // TODO: handle exception
    System.out.println("oo");
  }
}

/**
 * As the actions performed here requires Internet
     connection, they have to
 * be separated into a thread in order to avoid
     deadlocks.
 */
public void run() {
  this.currentCell = System.getProperty("com.
     sonyericsson.net.cellid") ;
  if(currentCell == null ){
    /*
     * this part is included to get the emulator
        to work during
     * development.
     */
    currentCell = "No Cell Available";
  }
  newCell(currentCell);

  try{
    // get the current cellId from the phone
    while(true){
      String newCell =
```

```
            System.getProperty("com.sonyericsson.net.
                cellid") ;
            if(newCell == null){
              newCell = "Error";
              System.out.print("Could not get cellID \n"
                  );
            }
            if(newCell.compareTo(currentCell) != 0 ){
              // we've moved in to a new cell and need
                  to update the page and the table
              newCell(newCell);
              currentCell = newCell ;
            }
            Thread.sleep(15000);
          }
        }
        catch (Exception e) {
          // TODO: handle exception
          e.printStackTrace();
        }
    }

    /**
     * This method is to be invoked every time the
         phone moves from one
     * cell to another. The method refreshes the list
         of visited cells,
     * and informs the web-server about its new
         location by calling
     * sentHttpGet().
     * @param cell
     */
    private void newCell(String cell){
      // move all the other cells one down
      for(int i = loc.length-1 ; i > 0 ; i--){
        loc[i] = loc[i-1];
      }
      // Create a new Position-object and put it at
          index[0]
```

```
Position now = new Position();
// put in the new one at index 0 of the list

loc[0] = now ;

// make a form and place all the text
page = new Form("Cells visited");
for(int i = 0 ; i < loc.length ; i++){
  if(loc[i] != null)
    page.append(loc[i].toString());
  else
    page.append("Not recorded");
}
page.addCommand(BACK_CMD);
page.addCommand(EXIT_CMD);
page.setCommandListener(new CommandListener(){

  public void commandAction(Command arg0,
     Displayable arg1) {
    if(arg0.equals(EXIT_CMD)){
      parent.notifyDestroyed();
    }
    else if(arg0.equals(BACK_CMD)){
      disp.setCurrent(parent.menuList);
    }

  }

});
// put the form in the displayable area'
disp.setCurrent(page);

// send a post-request to the tracing server

// prepare request that pushes our position-
   information onto the server....
String url = "http://fribyte.uib.no/~vidar/
   pusher/pushHandle.php?imei="+ imei
+ "&cid=" + now.getCid()
```

```
    + "&time=" + now.getTime().getTime()
    + "&lac=" + now.getLac()
    + "&mcc=" + now.getMcc()
    + "&mnc=" + now.getMnc() ;

    String resultStr ="";
    char tmpChar ;
    // Make sure the string is a valid url:
    for(int i = 0 ; i < url.length() ; i++){
      tmpChar = url.charAt( i );
      switch( tmpChar ) {
      case ' ' : resultStr+=( "+" );
      break;
      default : resultStr+=( tmpChar );
      break;
      }
    }
    // send the result to the server using a get
       request
    sendHttpGet(resultStr);

  }

  /**
   * Method that does an get request to a server.
   * In effect it pushes information about where the
      device is connected
   * to the network to the server.
   */
  private void sendHttpGet(String url) {
    HttpConnection httpConn = null ;

    try{
      httpConn = (HttpConnection) Connector.open(url
         ) ;
      httpConn.setRequestMethod(HttpConnection.GET);
      httpConn.setRequestProperty("User-Agent",
      "Profile/MIDP-1.0 Confirguration/CLDC-1.0");
```

```
        int respCode = httpConn.getResponseCode();
        /*
         * Check whether the request succeeded or not
            and let the
         * user know by showing an alert.
         */
        if(respCode == HttpConnection.HTTP_OK){
          Alert alert = new Alert("OK") ;

          alert.setTimeout(Alert.FOREVER);
          alert.setString("Pushed data");
          disp.setCurrent(alert);
        }
        else{
          Alert alert = new Alert("Error") ;

          alert.setTimeout(Alert.FOREVER);
          alert.setString("error");
          disp.setCurrent(alert);
        }
        httpConn.close() ;
      }
      catch (Exception e) {
        Alert alert = new Alert("Error") ;
        alert.setTimeout(Alert.FOREVER);
        alert.setString("There was an error pushing
           data");
        disp.setCurrent(alert);
        e.printStackTrace();
      }
    }
}
```

A.3.3 Geofetcher.java

This class is responsible for fetching the information needed for the "System Information" screen. It basically collects the information from the device using the API calls earlier described and displays them as an alert. In addition,

this class has implemented the functionality needed to fetch the latitude and longitude from my web server using the php script above.

```
package com.drageide.geolocalisation.CellIDPusher;

(Imports removed)

class GeoFetcher implements Runnable{
  // The address of the web-page recieving and
     handling the lookup of lat/lon
  String serverAdress = "http://www.fribyte.uib.no/~
     vidar/pusher/pushTest.php";

  String CID ;
  String MNC ;
  String MCC ;
  String LAC ;
  String location ;
  Display disp ;
  Alert al ;
  private GeolocalisationMidlet parentMidlet;

  /*
   * Constructor. Trivial.
   */
  public GeoFetcher(GeolocalisationMidlet parent){
    this.parentMidlet = parent ;
  }

  /*
   * Getters and setters removed as they contain
      nothing special
   */


  /**
   * Method that sends a get-request to a web-server
       who uses the parameters
```

```
     * to determine the apprixomate position of the
        mobile device.
     *
     * @return A string with the result of the query
        or throws an exception.
     */
    private String getFromUrl(){
      // Add the get-parameters to the request.
      serverAdress = serverAdress +"?cid=" + CID + "&
         lac=" + LAC + "&mcc="
        + MCC +"&mnc="+ MNC ;

      // Print the adress for debugging purposes
      System.err.println(serverAdress);

      // Set up connections and buffers
      StreamConnection sc = null ;
      InputStream ios = null;
      StringBuffer sBuff = new StringBuffer();



      try {
        sc = (StreamConnection) Connector.open(
           serverAdress);
        ios = sc.openInputStream();

        int ch ;
        while((ch= ios.read()) != -1 ){
          sBuff.append((char) ch);
        }
        return sBuff.toString();
      } catch (Exception e) {
        System.err.println("" +
        "Error while fetching data from web-server");
        return "" ;
      }
    }
```

```
  /**
   * Since this should be run in a seperate thread
     we need a run() method.
   *
   */
  public void run() {
    location = getFromUrl();
    al.setString(al.getString() + location );
    al.setCommandListener(parentMidlet);
    disp.setCurrent(al);

    try {
      Thread.sleep(5000);
    } catch (InterruptedException e) {
      // TODO Auto-generated catch block
      e.printStackTrace();
    }
    disp.setCurrent(parentMidlet.menuList);
    System.err.println(location);

  }
}
```

A.3.4 Position.java

This is a simple class for representing the position of a mobile device.

```
import java.util.Date;

/**
 *
 * @author Vidar Drageide
 *
 * This class implements the functionality needed
    for representing the
 * geographic position of a mobile device.
```

```
 * A position consists of information related to
    where the device is
 * connected to the network.
 * The class also contains fields that represent
    latitude and longitude,
 * as well as getters and setters for these fields.
 */
public class Position {
  // Network specific information
  private String cid;
  private String lac;
  private String mnc;
  private String mcc;

  // Geographic position holders
  double lat;
  double lon;

  private Date time;

  /**
   * Method that fetches information about the
      current connection to the
   * network from the phone.
   * The API used here is Sony Ericsson Specific and
      uses the
   * System.getProperty("") call.
   */
  public Position(){
    try{
      // Fetch Cell-ID from the phone
      this.cid = System.getProperty("com.sony
         Ericsson.net.cell ID");
      // Fetch Local Area Code (LAC)
      this.lac = System.getProperty("com.sony
         Ericsson.net.lac");
      // Fetch Mobile Country Code (MCC)
      this.mcc = System.getProperty("com.sony
         Ericsson.net.mcc");
```

```
      // Fetch Mobile Network Code
      this.mnc = System.getProperty("com.sony
         Ericsson.net.mnc");
      this.time = new Date();}
    catch (Exception e) {
      // TODO: handle exception
      System.err.println("An error occurred while
         fetching network" +
          " related data from the mobile device.\n")
             ;
    }
}

/*
 * Getters and setters, removed as none of them
    contain special
 * functionality.
 */


/**
 * Method for fetching a string containing date
    information on the format:
 * DDMMYYYY.
 * @return String with date in format DDMMYYYY
 */
private String getDDMMYYYY(){
  long a = time.getTime();
  a = a%1000 ;
  String dateString = time.toString() ;
  String year = dateString.substring(dateString.
     length()-4);
  String month = dateString.substring(4, 7);
  String date = dateString.toString().substring
     (8,10);
  String time = dateString.toString().substring
     (11,19);
  String parsedDate = year + "-" + date + "-" +
     month + " " + time + ":" + a ;
```

```
    return parsedDate ;
  }

  /**
   * Standard toString method.
   * @return String formatted: "Cell ID DDMMYYYY"
   */
  public String toString(){
    String ret = "";
    ret += cid + " " + this.getDDMMYYYY() + "\n" ;
    return ret ;
  }
}
```

Appendix B

Call Detail Records

I made an inquiry to my cell phone provider and asked them to provide me with a list containing my phone traffic. This list is included under as a figure for reference and completeness.

Chess *Communication AS*
Besøksadr. Møllendalsveien 1
Pb. 6142 Postterm. NO - 5892 BERGEN
Org.nr. NO 981 161 408

Dato	Kl.	Avsender	Mottaker	Varighet i sek.	Celle	Celle navn	Type
20090205	150358		4798637254	0	12011	BONTELABO	SMS
20090205	203455	4795909049	4798637254	12	54661	HEGRENESET	VOICE
20090206	132839	[redacted]	4798637254	100	12011	BONTELABO	VOICE
20090206	142310	[redacted]	4798637254	9	54661	HEGRENESET	VOICE
20090206	151820	[redacted]	4798637254	43	12011	BONTELABO	VOICE
20090206	175545	[redacted]	4798637254	0	42681	Ukjent	SMS
20090207	080124		4798637254	0	33086	FLESLAND	SMS
20090207	103858		4798637254	0	02251	GARDERMOEN	SMS
20090207	103902		4798637254	0	02251	GARDERMOEN	SMS
20090207	125640	[redacted]	4798637254	0	13138	PILESTREDET PARK	SMS
20090207	125642	[redacted]	4798637254	0	13138	PILESTREDET PARK	SMS
20090207	200519	[redacted]	4798637254	56	34731	MYRVOLL	VOICE
20090208	113903	[redacted]	4798637254	0	22683	Ukjent	1981
20090208	134532	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	141244	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	142937	[redacted]	4798637254	0	99999	Ukjent	MMS
20090208	142943	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	142946	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	143422	[redacted]	4798637254	0	99999	Ukjent	MMS
20090208	143429	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	143432	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	180808	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	180813	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	180826	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	180830	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	180834	[redacted]	4798637254	0	22683	Ukjent	SMS
20090208	212325		4798637254	0	22683	Ukjent	SMS
20090208	221354	[redacted]	4798637254	138	22683	Ukjent	VOICE
20090209	125319	[redacted]	4798637254	0	03758	RISLØKKA	SMS
20090209	151123	[redacted]	4798637254	0	03758	RISLØKKA	SMS
20090209	222130		4798637254	0	23668	GARDERMOEN TERMINAL	SMS
20090209	223723	[redacted]	4798637254	47	23919	GARDERMOEN PIR VEST	VOICE
20090210	074923	[redacted]	4798637254	0	54661	HEGRENESET	SMS
20090210	120142	[redacted]	4798637254	0	54661	HEGRENESET	SMS
20090210	133658	[redacted]	4798637254	0	54661	HEGRENESET	SMS
20090210	133946	[redacted]	4798637254	1	54661	HEGRENESET	VOICE
20090210	215420	[redacted]	4798637254	45	54661	HEGRENESET	VOICE
20090210	222101	[redacted]	4798637254	0	54661	HEGRENESET	SMS
20090211	141150	[redacted]	4798637254	0	54661	HEGRENESET	SMS
20090211	153401	[redacted]	4798637254	56	12011	BONTELABO	VOICE
20090211	164005	[redacted]	4798637254	0	54322	GAMLE BT	SMS
20090211	165621	[redacted]	4798637254	38	12028	MØHLENPRIS	VOICE
20090211	201703	[redacted]	4798637254	270	12011	BONTELABO	VOICE
20090211	223747	[redacted]	4798637254	154	12011	BONTELABO	VOICE
20090211	225422	[redacted]	4798637254	22	12011	BONTELABO	VOICE

Chess *Communication AS*
Besøksadr. Møllendalsveien 1
Pb. 6142 Postterm. NO - 5892 BERGEN
Org.nr. NO 981 161 408

Figure B.1: A scan of the document sent to me by my provider. Most of the originating numbers are blurred out for privacy reasons.