REPORTS
IN
INFORMATICS

ISSN 0333-3590

Personal Information Leakage:
A Study of Online Systems in Norway

André N. Klingsheim and Kjell J. Hole

REPORT NO 370 February 2008

Department of Informatics
UNIVERSITY OF BERGEN
Bergen, Norway

This report has URL
http://www.ii.uib.no/publikasjoner/texrap/pdf/2008-370.pdf

Reports in Informatics from Department of Informatics, University of Bergen, Norway, is available at http://www.ii.uib.no/publikasjoner/texrap/.

Requests for paper copies of this report can be sent to:

Department of Informatics, University of Bergen, Høyteknologisenteret,
P.O. Box 7800, N-5020 Bergen, Norway

Personal Information Leakage: A Study of Online Systems in Norway[*]

André N. Klingsheim and Kjell J. Hole

NoWires Research Group
Department of Informatics
University of Bergen, Norway
Email: {klings,kjellh}@ii.uib.no

Version 1.4, February 14, 2008

Abstract

Governments and commercial companies connect more and more computer systems to the Internet, giving people easier access to services. Many of these online services handle personal information. Leakage of such information can facilitate large-scale identity theft. This report determines how personal information leaks from online systems of national importance, discusses proof of concept software to demonstrate the seriousness of the problem, and suggests how to improve the situation.

1 Introduction

Governments have traditionally provided public services to citizens through various offices, where citizens wait in line to speak with government representatives. Modern governments provide a better alternative by making many of their services available on the Internet. These online services are referred to as e-Government. Citizens save time because they avoid lengthy queues and are freed from various offices' opening hours. The government also saves resources, as citizens to a larger degree can help themselves without talking to government employees. Similarly to e-Government, many companies offer services to their customers on the Internet. These online services often manage personal information. Some are even required by law to collect such information. The services discussed in this report are available on the Web, and users access them through their Web browser.

Adequate security and privacy are vital for online systems containing personal data. *Information privacy* refers to the individual's interest in controlling the flow of personal information [2, p. 63]. It can be difficult for a citizen to keep track of what information is available where, and to whom on the Internet. For example, a citizen might have an account in a governmental service without even knowing it.

Identity theft occurs when someone uses another individual's personal information to pose as that individual [2, p. 99]. Useful information is e.g. credit card numbers and expiration dates, usernames/passwords, date of birth, identification numbers, name, and address of a victim. Successful impersonation of a victim lets the identity thief commit fraud.

Frequently mentioned threats to an individual's privacy are dumpster diving, phishing, pharming, trojan horses, and hacking. In addition, there are major privacy concerns related

[*] Earlier versions of this report were entitled: "Identity Theft: Much too Easy? A Study of Online Systems in Norway". A short version of the report is published in the proceedings of the Financial Cryptography and Data Security Conference 2008 [1].

to websites used to maintain social networks, such as myspace.com or facebook.com, where people voluntarily share personal information. An individual can try to protect his privacy by being careful about giving out personal information both online and offline, and keeping his operating system, firewall, and anti-virus software up to date.

However, a major problem is the information beyond the individual's control, which the individual cannot secure [3]. Large amounts of data leak from various systems, and governments seem to be struggling the most to keep the data safe [4], [5, p. 28]. During 2006, there were several news stories in Norway where various governmental institutions disclosed personal information on the Internet by accident. However, the amount of leaked information was insignificant compared to the scenarios described later in this report.

Norwegian Birth Numbers (NBNs, Norwegian: fødselsnummer) are in widespread use in national computer systems in Norway. The NBNs are National Identification Numbers (NINs) comparable to the American Social Security numbers. Many countries have NINs, see [6] for pointers to governmental websites with information about NINs.

NBNs have been used as tokens of authentication by governmental institutions and companies in the private sector since long before the age of online services. This solution has worked well, still, identity theft has been possible with knowledge of a person's NBN and name for a long time. Because NBNs are still widely used as authenticators, they are of great value to an identity thief. The Norwegian Data Inspectorate has expressed concern over the use of NBNs as usernames in e.g. online banking systems. The problematic use of NBNs by a Norwegian pension fund was described in [7].

This report discusses privacy issues related to the use of NBNs in online Norwegian systems, but should be relevant to other countries using NINs as well. Steps on how to prepare large-scale identity theft are outlined, and proof of concept software automating the collection of personal information is described. Major privacy violations are highlighted and steps to reduce the problem are suggested.

The rest of this report is organized as follows. Section 2 discusses personal identifiers and how they are used, Section 3 determines why systems reveal personal information, and Section 4 explains how a large-scale identity theft can be prepared. Section 5 then highlights how economic damage can be inflicted through misuse of online systems, Section 6 makes general suggestions on how to improve the current situation, and Section 7 concludes the report.

2 Identifiers and identities

An *identifier*, such as a name, NIN, or a customer number, points to an identity. The *identity* of an individual is the set of information associated with that individual in a particular computer system [2, p. 20].

Identifiers should be chosen with great care when designing a system. Certain identifiers can make the task easier for those who want to collect information about individuals. Structured identifiers should only be used after careful analysis of possible side effects, such as distributed denial of service attacks on online banks [8].

Identifiers have different scope. Some are global, e.g. e-mail addresses, some are national, like NINs, while others have a rather limited scope, such as a customer number only meaningful to a specific company. The scope of an identifier used in a system affects the users' privacy. In national computer systems in Norway, a user can seldom choose her identifier and is often forced to use her NBN. By design, systems then facilitate the creation of personal profiles, since an NBN refers to an identity which can be directly linked to a unique individual.

The Norwegian Birth Numbers consist of 11 digits: $x_1x_2x_3x_4x_5x_6i_1i_2i_3c_1c_2$	
$x_1x_2x_3x_4x_5x_6$	**Birth date** (ddmmyy).
$i_1i_2i_3$	**Individual number**. Highest available number for that day is used for each person. For persons born in 1854–1899 the possible numbers are 500–749, anyone born between 1900 and 1999 are given an individual number in the range 000–499. For people born 1940–1999, the secondary range 900–999 is also used. Finally, for those who are born in 2000–2039 the numbers 500–999 are used. Girls have an even i_3 value while boys have an odd i_3 value.
c_1c_2	**Control digits**. Used to detect common human errors when people supply an NBN, such as one digit wrong or any two digits exchanged. $c_1 = 11 - ((3x_1 + 7x_2 + 6x_3 + x_4 + 8x_5 + 9x_6 + 4i_1 + 5i_2 + 2i_3) \bmod 11)$ $c_2 = 11-((5x_1+4x_2+3x_3+2x_4+7x_5+6x_6+5i_1+4i_2+3i_3+2c_1) \bmod 11)$ If c_1 or c_2 is 10, then the number is rejected and the next possible number is chosen. If c_1 or c_2 is 11, then 0 is used instead.

Table 1: Structure of Norwegian Birth Numbers

2.1 The Norwegian Birth Number

All Norwegian citizens are assigned an NBN, containing the date of birth and reflecting the gender of an individual, as shown in Table 1 [9, 10]. NBNs are assigned chronologically for a particular day, yielding a sub-range of used NBNs within the range of all valid NBNs for that day. NBNs are not secret by Norwegian law, but access to them is restricted.

Let the set of all possible correct NBNs according to Table 1 be denoted by C. This set can easily be generated by a computer. Furthermore, let the set of correct NBNs that are assigned to individuals be denoted by A. This set is predictable since the individual numbers from Table 1 are assigned chronologically. Note that the number of NBNs assigned to people born on a particular day can increase over time as NBNs are assigned to individuals with residence permit for a minimum of 6 months. Finally, let U represent the set of assigned NBNs used in a particular system. This yields the following relations: $U \subseteq A \subseteq C$.

An identity thief can generate C offline, and is interested in determining U for selected systems. Establishing A will make it easier to determine U and is a preliminary step to prepare large-scale identity theft.

Figure 1 depicts different NBN sets, using a mobile operator, an online bank, and a university as examples. The Norwegian legal age is 18, therefore, NBNs for all Norwegian citizens aged $\geq$ 18 years, denoted U_{mo}, are valid in the mobile operator's signup system. For an identity thief, it is valuable to determine U_{mo} since this set contains all NBNs that can be used to commit financial fraud. Furthermore, U_{mo} can be used to determine users in other systems such as the online bank or the university, denoted U_{ob} and U_u in Figure 1.

2.2 The National Identity Register

The Norwegian National Identity Register (NNIR) (Norwegian: Folkeregisteret) contains the NBN, full name, full address, place of birth, and family relations for all Norwegian citizens. Approximately 7 million identities are kept in the registry, where 4.5 million people are residents in Norway and the rest are emigrants. The NNIR is often used to determine full name and address of an individual. Certain requirements defined by Norwegian law must be fulfilled to be allowed to interact with the NNIR. The Office of the National Registrar (Norwegian: Sentralkontoret for folkeregistrering) grants applicants access to the registry. Two levels of access exist.

Public authorities can apply for full access to the NNIR, but they have to document a need to be granted access. Authorities with full access can perform searches using NBNs, names, or addresses as search keys. Lists of NBNs and corresponding names can be extracted from the registry.

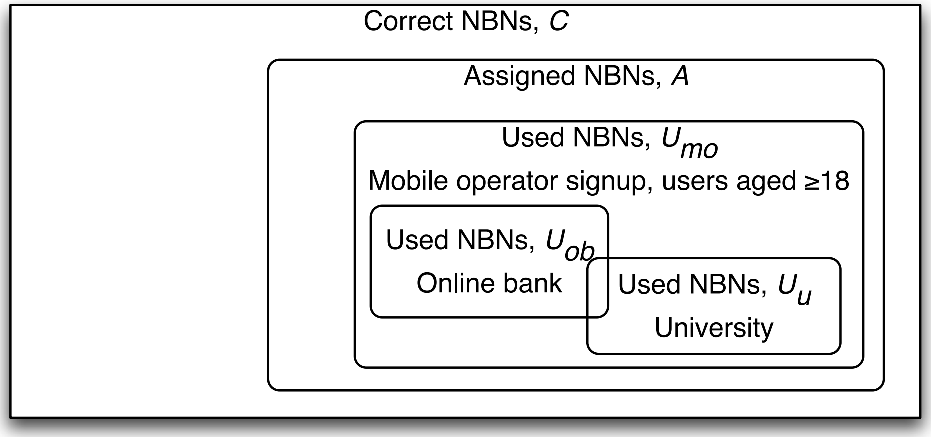

Figure 1: NBN classification

Companies in the private sector and public authorities can apply for limited access, providing less sophisticated search opportunities. Only “exact search” is possible, where the minimum search input is the date of birth and name of an individual. A query will then return a single identity with the NBN and address of an individual.

Many governmental and commercial entities use information from the registry. In a 2005 press release, Skatteetaten stated that about 1 500 entities had access to the registry, and 30 million queries were executed. The NNIR and the entities accessing the registry constitute a large *national identity system*, with the NNIR holding the base identities and users of the registry holding personal profiles extending the base identities. A report from the National Research Council in the US raises important privacy questions regarding national identity systems [11]. In particular, the individuals with identities in a system should know what information is stored, how it is made available (to any third party), be able to keep the information updated, and most importantly, be able to prevent disclosure of information. These requirements are not fulfilled in the Norwegian identity system based on the NNIR.

3 Why systems leak

Individual authentication, referred to as authentication for simplicity, is the process of establishing an understood level of confidence that an identifier refers to a specific individual [2, p. 19]. The authentication is said to be strong if the level of confidence is high. Generally, online services authenticate users before authorizing access to personal information.

The privacy issues highlighted in this report are caused by the use of NBNs as identifiers, and one or more of the following: privacy violating authentication schemes, weak authentication schemes, and lack of authentication.

3.1 Username/password authentication

Many websites reveal identifiers’ validity because of their authentication schemes. Figure 2 illustrates a popular solution in Norwegian systems where a user first enters his NBN, the system verifies that the NBN is used, and then asks for authentication information such as a password or Personal Identification Number (PIN). A software program can post candidate NBNs to such a website and log which NBNs are used. Hence, the complete set U of used identifiers can easily be determined. Online services in this category include

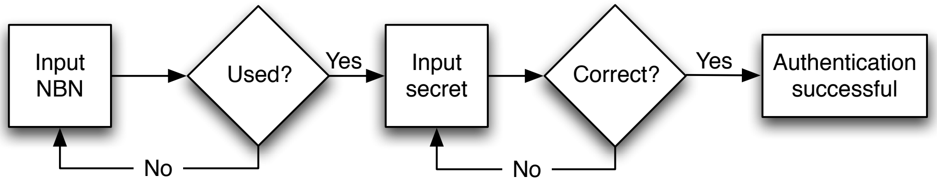

Figure 2: Privacy violating authentication scheme

e.g. governmental websites, online banks, and student portals at several universities and colleges.

3.2 Weak authentication

Unfortunately, many online systems employ weak authentication of users. One example is a service where citizens can change their primary physician. A user enters his NBN, name, and zip code to log into the system, all of which are available on other websites. The user's assigned physician is revealed after a successful login procedure. The system enforces a limit of two changes of primary physician per year. An evil minded person would of course log in and do this change twice for a victim.

Even worse, during 2007 several mobile operators leaked names and addresses corresponding to NBNs during their signup process, effectively publishing data from the NNIR on the Internet. Users selected a subscription type and entered their NBN to sign up. The mobile operator then conveniently presented the full name and address associated with the NBN on the webpage for user confirmation. Since an NBN and a name suffice as authenticators in many online and offline systems, an identity thief could use these web services as a starting point before targeting other systems.

3.3 No authentication

Systems exist that do not authenticate users, but still reveal information. In particular, one online registry in Norway lets users check prescription rights for medical personnel. The database is searchable using NBNs as the search key. Criminals or drug addicts can easily build a list of physicians with prescription rights, and attempt to bribe, con, or pressure them to issue prescriptions.

4 Preparing for identity theft

This section outlines the steps needed to prepare large-scale identity theft, and describes software developed to demonstrate how easily personal information can be obtained.

4.1 Choose the victims

Potential identity theft victims can be selected by age, gender, a geographical area, and/or presence in a particular system. NBNs can be generated for a certain age group, down to a specific day. Users' presence in a particular system and their residence can be established through privacy violating services.

4.2 Harvest identities

Late summer 2007 the previously mentioned mobile operators' websites were exploited by unknown parties, and a large number of identities were downloaded. The media reports a total of 140 000 identities in the wild. The accuracy of this number is undetermined, but it is reasonable to believe that this number is too low.

Based on the identities leaked, more information can be extracted from other systems in order to create personal profiles of citizens. However, the knowledge of a person's name, address, and NBN is in many cases enough to commit identity theft.

4.3 Establish NBN validity in systems

At least two Norwegian banks leak their customer lists due to their authentication procedures, resembling the process in Figure 2. Combining the list of bank customers' NBNs with the identities retrieved from the mobile operators, enables identity theft against bank customers. As an example, to cancel a customer's payment cards it is sufficient for an attacker to call the bank and give the customer's NBN and name. The bank will issue new cards, delivered to the customer by mail. The attacker can then steal these cards from the customer's mailbox.

4.4 Harvest other useful information

Not all information is directly linked to an NBN, but it is related to a name or an address. In Norway, several companies run registries with phone numbers, names, and addresses. Multiple registries are accessible on the web, providing search opportunities. Results can be obtained for a specific name, a specific address, or a specific phone number. Most Norwegians with a stationary phone or a cell phone subscription are included in the registries. A phone user can choose to not be in the registries, but has to pay for this service. Therefore, the majority of subscribers are included in the registries.

Many phone subscriptions paid by employers are registered with the name or address of the company and the name of the employee. This information is included in the registries, hence, the employer can be determined for many people.

4.5 The software

To establish how easy it is to automate harvesting of personal data, the first author developed a small graphical program in the Java programming language, called *NBNtool*. The software is discussed in more technical detail in the appendix. Two notable features are discussed here.

NBNtool was able to establish many of the customers in one of the largest banks in Norway, taking advantage of the bank's authentication scheme resembling Figure 2. The bank offered three authentication tokens, a PIN card, and two PIN generators, denoted PG1 and PG2. A customer entered his NBN and was redirected to a login page for his particular authentication token. Users with PG1 or a PIN card were easily determined. However, entering the NBN of a non-customer led to the login page for PG2. Hence, to distinguish between customers with PG2 and non-customers, a login attempt with a PIN was required. The authors did not want to carry out this step because it could be viewed as an attempt to break into accounts. NBNtool therefore only revealed customers with PG1 or a PIN card.

Furthermore, NBNtool used a particular mobile operator's signup procedure to extract full name and address for Norwegian citizens aged ≥ 18, by simply posting NBNs to the website. Hence, large parts of the NNIR could be mirrored through the mobile operator's website.

NBNtool communicated with the websites through The onion routing (Tor) network to avoid detection [12]. The bank is known to utilize intrusion detection technology, but NBNtool still ran uninterrupted on several occasions.

4.6 Sharing the information

Software automating the exchange and verification of credit card numbers online has existed for many years [13]. It is also well known that personal information is traded on various websites. According to a newly released report, 15% of the underground servers used to trade personal information are located in Sweden [5, p. 30]. There is no reason to believe that Norwegians' personal information is not already on the market. Sweden is after all our closest neighbor.

5 Economic effects for system owners

It is trivial to inflict economic loss through online services that use NBNs and names as authenticators. Several mobile operators check the creditworthiness of potential customers before the signup process is completed. The mobile operators buy this service from a third party. If an identity thief executes incomplete sign-up processes for a large amount of individuals, the mobile operators will have to pay for many worthless credit checks. In addition, creditscore companies must send a letter to each individual stating that a credit check took place on behalf of the mobile operators. Since most people are reluctant to give out their financial status for no reason, the incident will most certainly be covered by the national news and lead to substantial damage to the mobile operators' reputation.

The government provides several services where citizens can order information or forms online. Often, it is sufficient to enter an NBN, and the information or form is mailed to the corresponding address in the NNIR. Again, it is easy to cause economic loss by ordering information or forms for a large number of NBNs.

A slightly stronger effort is needed to affect systems relying on NBN/password authentication. A number of unsuccessful login attempts must be carried out to shut down an account. For example, some online banks, used by many companies and individuals, can be virtually shut down by a distributed denial-of-service attack on the application layer [8]. The economic consequences would be severe.

6 Improving the situation

There is no quick solution to the current problem of information leakage. NBNs have been misused as secrets, or authenticators, for years. Slowly the threat model has changed, as NBNs are used in more and more systems connected to the Internet. The NNIR based identity system has seen a continuos growth, and now lacks clearly defined boundaries and areas of use. According to [11], an identity system should undergo a thorough analysis involving all stakeholders. Both the creators of the system and the users must be involved in the analysis. Scientists with expertise on privacy, and without commercial interests in the system, should also partake to ensure that citizens' privacy is well protected.

A proper analysis would take years to complete and be a highly challenging task. However, the current NBN system is only functional to 2039. A new identity system, adapted to current and future challenges, should be developed. The work should start now. Hopefully, a significantly better and more modern system will see the light of day long before 2040.

Still, measures are needed to reduce the problems seen today. The gross privacy violations on several websites show that website owners lack incentives for protecting users' privacy. This has to change, as the consequences for victims of identity theft can be dire. Authorities responsible for privacy in Norway need to find better ways to work in the future, enabling them to deal with privacy violators in an efficient and swift manner. The findings described in this report clearly show that the control of personal information is unsatisfactory. The authorities' shortcomings in this area have many explanations, including judicial limitations, lack of funding, shortage of staff, and unclear placement of liability.

An important change that would improve the current situation is to enforce regulations on services with privacy implications so that users have to opt-in to access the service online. Today, individuals have to locate privacy violating services and try to invalidate their identities in these services. However, it may be impossible for the user to opt-out. An example is an online service provided by Norway Post, the company handling all mail in Norway. A victim of identity theft discovered that her mail was rerouted and that credit cards were ordered in her name. After spending several days to regain control of her mail, the identity thief misused the online service again. Spending only a few minutes online, using the victim's NBN and name, the perpetrator rerouted her mail once more. Norway Post stated that their system did not support shutting down access for particular users, leaving the lady in a rather unfortunate situation.

7 Conclusions

Data harvesting is possible in Norwegian online systems. Large amounts of NBNs and corresponding personal information can be determined. Many websites use NBNs to identify, or even authenticate their users, facilitating creation of personal profiles. We conclude that large-scale identity theft is indeed possible in Norway. The risk of this happening is unclear, but it is definitely present as small-scale online identity theft is already a problem.

Both government and industry make it easy to determine identifiers used in several of their online systems. Furthermore, the government forces companies such as mobile operators to collect NBNs, names, and addresses from their customers and compare them to the records in the NNIR, but does not adequately ensure that the users' privacy is protected in the process.

NBNs should not be used as authenticators anymore. They are in practice published on the Internet and can easily be collected. In addition, there are probably thousands of people with authorization to access the NNIR. NBNs must therefore be considered public information in the future.

New authentication schemes must be selected and put to use in Norwegian online systems. The strength of the authentication should reflect the sensitivity of data stored in a system, and privacy aspects of the authentication scheme itself should be considered.

This report highlights severe privacy issues, but the whole picture cannot be analyzed in a single report. A thorough analysis of the current NBN-based identity system in Norway is called for, and will lay the groundwork for the development of a new and improved identity system that needs to be in place before 2055.

7.1 Final remarks

Due to the potentially severe consequences of the current privacy violations in Norway, the authors demonstrated NBNtool for the Norwegian Data Inspectorate and the Financial Supervisory Authority of Norway in early January, 2007. NBNtool worked with the mobile operator's signup service without modification until August 2007. The online bank had then made minor changes, but NBNtool was easily adapted to the new application protocol.

All privacy violating websites found during the project were reported to the Data Inspectorate in January 2007. In February 2007, the online bank was notified directly since their login procedure still leaked parts of their customer list. The bank replied that the issue had been taken into account during the risk analysis of their system.

The Data Inspectorate has informed us that there are ongoing inspections of the privacy violating websites discussed in this report. Furthermore, the Inspectorate has an ongoing project related to identity theft and the use of NBNs. Finally, several future inspections will focus on the use of NBNs and information leaks.

In Norway, science projects where personal information is handled electronically must be approved prior to project startup. It is somewhat ironic that the authors had to apply to the Norwegian Ombudsman for Privacy in Research to get permission to electronically handle personal information that is floating more or less freely on the Internet.

7.2 Acknowledgments

We thank the Office of the National Registrar for valuable feedback on Section 2.1, and for taking the time to answer our questions during the authoring of Section 2.2.

Special thanks are due to the Norwegian Data Inspectorate for allowing us to demonstrate NBNtool at a meeting in January of 2007. The first author is also very grateful to Senior Engineer Atle Årnes for several useful discussions on privacy issues.

Appendix: NBNtool

This appendix describes some of the technicalities of NBNtool, and proxy software used to inspect privacy violating websites.

Using off-the-shelf communication libraries the development of NBNtool was simplified. The Jakarta Commons HttpClient library was used to perform the communication with the websites [14]. This library saves developers time by handling cookies and redirections. Also, HTTP requests are easily created and executed, and HTTP responses are returned as convenient data structures. An important feature is the ability to communicate through a web proxy, such as the default frontend to the Tor network client, Privoxy [15]. Minor configuration changes enabled NBNtool to communicate through Tor.

The WebScarab proxy software was used to inspect the communication between the web browser and the websites [16]. A developer can configure his web browser to communicate through the proxy, and record a session with a website. Afterwards, the entire session can be studied to determine cookies and parameters used throughout the communication.

To create a tool such as NBNtool a developer needs a basic understanding of the HTTP protocol, and must be capable of basic network and GUI programming. Thousands of developers do this for a living, and many universities and colleges have courses on these subjects in their computer science departments.

References

[1] A. N. Klingsheim and K. J. Hole, "Identity Theft: Much too Easy? A Study of Online Systems in Norway," in *Proc. Financial Cryptography and Data Security*, January 2008. ∗

[2] S. T. Kent and L. I. Millett, editors, *Who Goes There? Authentication Through the Lens of Privacy*, The National Academies Press, 2003. 1, 2, 3

[3] B. Schneier, "Risks of Third-Party Data," *Communications of the ACM*, 48(5):p. 136, May 2005. 1

[4] Privacy Rights Clearinghouse, "A Chronology of Data Breaches," `http://www.privacyrights.org/ar/ChronDataBreaches.htm`, last checked Feb. 14, 2008. 1

[5] Symantec Inc., "Symantec Internet Security Threat Report XI," March 2007. 1, 4.6

[6] Wikipedia, "National identification number," `http://en.wikipedia.org/wiki/National_identification_number`, last checked Feb. 14, 2008. 1

[7] V. Moen, A. N. Klingsheim, K. I. F. Simonsen, and K. J. Hole, "Vulnerabilities in E-Governments," *International Journal of Electronic Security and Digital Forensics*, 1(1):pp. 89–100, 2007. 1

[8] K. J. Hole, V. Moen, and T. Tjøstheim, "Case Study: Online Banking Security," *IEEE Security & Privacy*, 4(2):pp. 14–20, March/April 2006. 2, 5

[9] E. S. Selmer, "Personnummerering i Norge: Litt Anvendt Tallteori og Psykologi," *Nordisk Matematisk Tidsskrift*, 12:pp. 36–44, 1964, in Norwegian. 2.1

[10] Skatteetaten, "Generelt om folkeregistrering," `http://www.skatteetaten.no/Templates/Artikkel.aspx?id=6640`, in Norwegian, last checked Feb. 14, 2008,. 2.1

[11] S. T. Kent and L. I. Millett, editors, *IDs—Not That Easy: Questions About Nationwide Identity Systems*, The National Academies Press, 2002. 2.2, 6

[12] Tor, "Anonymity online," `http://tor.eff.org`, last checked Feb. 14, 2008. 4.5

[13] B. McCarty, "Automated Identity Theft," *IEEE Security & Privacy*, 1(5):pp. 89–92, September/October 2003. 4.6

[14] The Apache Jakarta Project, "Jakarta Commons HttpClient," `http://hc.apache.org/httpclient-3.x/`, last checked Feb. 14, 2008. 7.2

[15] Privoxy, "Privoxy," `http://www.privoxy.org`, last checked Feb. 14, 2008. 7.2

[16] Open Web Application Security Project, "WebScarab," `http://www.owasp.org/index.php/Category:OWASP_WebScarab_Project`, last checked Feb. 14, 2008. 7.2