telektronikk 2.07

Privacy in Telecommunications

Telektronikk
Volume 103 No. 2 – 2007
ISSN 0085-7130

Editor:
Per Hjalmar Lehne
(+47) 916 94 909
per-hjalmar.lehne@telenor.com

Editorial assistant:
Gunhild Luke
(+47) 415 14 125
gunhild.luke@telenor.com

Editorial office:
Telenor R&I
NO-1331 Fornebu
Norway
(+47) 810 77 000
telektronikk@telenor.com
www.telektronikk.com

Graphic design:
Design Consult AS (Odd Andersen), Oslo

Layout and illustrations:
Gunhild Luke and Åse Aardal,
Telenor R&I

Prepress and printing:
Rolf Ottesen Grafisk Produksjon, Oslo

Circulation:
3,900

Networks on networks

Connecting entities through networks – in technological, societal and personal terms – enables telecommunication. Networks occur on different levels, form parts of larger networks, and exist in numerous varieties. The artist Odd Andersen visualises the networks on networks by drawing interconnected lines with different widths. Curved connections disturb the order and show that networks are not regular but are adapted to the communication needs.

Per H. Lehne, Editor in Chief

Contents

Guest Editorial

GEIR M. KØIEN, VLADIMIR A. OLESHCHUK

Geir M. Køien is a researcher in the Network Technologies group of Telenor R&I

Vladimir A. Oleshchuk is Professor of Computer Science at Agder University College, Norway

It was the best of times, it was the worst of times,
it was the age of wisdom, it was the age of foolishness,
it was the epoch of belief, it was the epoch of incredulity ...

– Opening lines of Charles Dickens, *A Tale of Two Cities*

The modern society is rapidly becoming a digital society. Our everyday activities are increasingly being registered and recorded. The registrations include data concerning our buying habits, medical records, reading- and viewing preferences, and travelling habits. The collected data is recorded and analyzed in detail. The growing number of personal devices and gadgets that we use emphasises the scale of the digitalizing of our lives. Suffice to mentions the ubiquitous mobile phone, iPods (and other music players) and digital cameras. These gadgets, being personal, tend to contain a lot of privacy sensitive data.

Our personal privacy is threatened. The threat does not so much come from a 1984 style Big Brother, but rather from a set of smaller big brothers. The small big brothers are companies that we interact with; they are public services and institutions. Many of these little big brothers are indeed also being invited to our private data by ourselves. By giving up a little privacy we may stand to gain. There are benefits in the form of convenience and improvement of personal safety, security and quality of provided services. The benefits can also transcend the individual and be of benefit to society if one is able use the accumulated private information wisely. Examples could include discovering new treatment methods or discovering disease outbreaks by mining patient medical records, monitoring of dementia and surveillance of private houses and apartments to improve personal safety and security.

Technology is an indiscriminate tool. With selective informed use of available technology it is still possible to remain almost completely anonymous in our society. It is also possible to avoid all eavesdropping by use of cryptographic technologies. Many of the necessary technologies are already available. However, the same technologies that provide you with your privacy can also be used to hide hideous crimes and terrorist activity. Suffice to mention drug trafficking, human trafficking, child pornography, terrorist activity and other despicable activities.

Society must protect itself against these activities, for unless it is able to protect itself it will break down. And we collectively cannot afford that to happen.

The traditional view is that many of the new services will invade people's privacy and there clearly is a trade-off between safety/security and privacy in the sense that we may have to give up some personal privacy in exchange for better safety/security. Here, like in so many other circumstances, there is an inherent asymmetry in who stands to gain and who must give the most. Companies and large public institutions don't take the individual's stand unless there either is something to be gained (reputation, customer satisfaction etc), or they are forced to respect our privacy due to regulations and laws.

Personal privacy as a subject can be problematic. At the extreme it is personal freedom against safety and security. There are also other tradeoffs like finding a balance between surrendering our privacy vs. benefiting in terms of improved user convenience. We shall not take a political stand on personal privacy and what level of personal freedom and privacy is the correct one. Our goal with this feature edition of *Telektronikk* is to point out and highlight some of the personal privacy problem areas and to provide the reader with pointers to further reading. Then we may have an informed debate on this fascinating and important topic.

The articles in this edition do not cover all aspects of personal privacy; that would be an impossible task for such a fast growing area of research. However, we hope that we have collected a balanced set of articles that gives the reader a reasonable overview of the topic.

The articles are loosely organized under the headings **Privacy Overview**, **Regulatory Aspects**, **Communication Privacy** and **Other Personal Privacy Areas**.

The first article, *Personal Privacy in a Digital World*, is an introductory overview article by the feature

 ISSN 0085-7130

editors. The article looks at many of the privacy threats and risks that we all must face in the modern society.

The second article, *Secure Multi-party Computations and Privacy Preservation: Results and Open Problems*, is written by one of the feature editors and Vladimir Zadorozhny, and is a relatively technical article on the use of secure multi-party computations (SMC) to provide privacy preserving properties. The mathematics behind SMC is intriguing, but also quite counterintuitive unless one is familiar with concepts like homomorphic public-key crypto-systems. In the article the authors try to demonstrate how we can design services that traditionally require customers' private data without giving away such private data.

In the **Regulatory Aspects** section there are three main articles covering amongst other things the lawful interception (LI) and data retention directive (DRD) requirements. The first article, *Privacy and Protection of Personal Data*, is written by Kjetil Rognsvåg who is the Privacy Ombudsman for Telenor ASA. The second article, *The EU Directive on Data Retention – An End to Justify the Means*, is written by Berit Svendsen who is currently a vice-president in Telenor Nordic Fixed. The third main article, *Lawful Interception*, is written by Rupert Thorogood and Charles Brookson, both involved in the standardization of LI specifications within Europe. Finally, the section ends with a short cautionary article, *Privacy and the Regulatory Big Brothers*, by one of the feature editors.

The third section, **Communication Privacy**, contains three relatively technical articles. The first one, *Subscriber Privacy in Cellular Systems*, is written by one of the feature editors and discusses the problems of location- and identity privacy in today's cellular networks. The second article, *Location Hidden Services and Valet Nodes*, by Lasse Øverlier and Paul Syverson, is not directly focused on personal privacy, but discusses privacy from a network perspective.

The final article in the section is an article on privacy in the context of wireless sensor network; the article is aptly entitled *A Framework for Efficient Security and Privacy Solutions in Data Intensive Wireless Sensor Networks* by Vladimir Zadorozhny et al.

The last section is **Other Personal Privacy Areas**. This section contains three articles. The first article, *RFID and Privacy*, is written by one of the feature editors and is an overview article on the exciting and sometime scary area of personal privacy vs. pervasive and possibly intrusive use of Radio-frequency Identification (RFID) technology. The second article in this section, *Privacy Preserving Data Mining in Telecommunication Services*, is a technical article written by Ole-Christian Granmo and one of the feature editors. The article looks at the possibility of permitting use of privacy sensitive data in data mining efforts while still preserving the personal privacy of individuals. The final article, *Distributed Health Records, Cryptographic Pseudonyms, and Privacy*, is written by Stig F. Mjølsnes and covers a very important area. Few things are more personal and potentially more sensitive than your health record, and given that the digital revolution is now clearly coming to the healthcare sector it certainly is important that the patient's personal privacy right is fully respected.

We hope you enjoy this issue of *Telektronikk* and we hope that it will contribute to keeping personal privacy on the agenda. Only then can we make informed decisions and find a reasonable balance between personal privacy and democratic values on the one hand and the promise of improved safety, security and everyday convenience from the emerging technologies on the other.

Geir M. Køien is a researcher in the Network Technologies research group of Telenor R&I. He holds a BSc hons. in Computing Science from University of Newcastle upon Tyne (UK), an MSc in IT from the Norwegian University of Science and Technology (NTNU), and is expecting to defend his PhD at the University of Aalborg (Denmark) in the autumn of 2007. Geir Køien has been working with mobile communication technology, including NMT, GSM/GPRS and UMTS, and system security aspects for the past 15 years. During the last few years his research focus has included privacy aspects. He has also been the Telenor delegate to the 3GPP SA3 (Security) workgroup since 1999.

email: geir-myrdahl.koien@telenor.com

Vladimir A. Oleshchuk is Professor of Computer Science at Agder University College. He received his MSc in Applied Mathematics (1981) and PhD in Computer Science (1988) from the Taras Shevchenko University in Kiev, Ukraine, and his MSc in Innovations and Entrepreneurship (2007) from the Norwegian University of Science and Technology (NTNU). From 1987 to 1991 he was Assistant Professor and then Associate Professor at the Taras Shevchenko University. He has been working at Agder University College since 1992. His current research interests include formal methods and information security with special focus on telecommunication systems.

email: vladimir.oleshchuk@hia.no

Personal Privacy in a Digital World

GEIR M. KØIEN, VLADIMIR A. OLESHCHUK

Geir M. Køien is a researcher in the Network Technologies group of Telenor R&I

Vladimir A. Oleshchuk is Professor of Computer Science at Agder University College, Norway

"I have as much privacy as a goldfish in a bowl."
– attributed to Princess Margaret of England

Most of us don't have to fear ending up with as little privacy as Princess Margaret, but there is nevertheless reason to defend our privacy vigilantly. In the following we shall examine the concept of personal privacy and take a look at some of the real-world challenges and threats that exist.

1 Introduction

In this article we present an overview of some personal privacy areas and of some of the threats to your personal privacy in a digital world. The scope is not to cover every possible angle and every possible personal privacy aspect, but to illustrate the breadth and pervasiveness of the personal privacy issues.

The article is organized in the following main sections:

- Personal Privacy Areas
- Real-World Privacy Threats
- User Consent is no Silver bullet
- Where Personal Privacy Ends

2 Personal Privacy Areas

2.1 Data Privacy

The *Data Privacy* property can be seen as an access restriction in the sense of "protection from intrusion and information gathering" [1]. That is, a *Data Privacy* requirement means that the visibility of the object shall be restricted. Thus, in terms of a file system this would mean that the file should have restrictions on the *read* rights. That is, only people or processes[1)] with the corresponding *read* permission is allowed to see the object.

Similarly, for a data object (message) in transit one must ensure that only those entities with the proper permission are allowed to read the object off the channel. The nature of the communication channels differ substantially; some channels may be able to provide physical protection while other channels, like radio channels, are easily read by any entity in the channels' proximity. To protect these channels one has to turn to cryptographic techniques like encryption.

So how strict is the privacy requirement? The question is relevant since some of the protection schemes are costly to operate and may cause considerable inconvenience to the users. The question is often best answered in terms of how valuable it is for you to keep the object private and who do you realistically see as your 'enemy'. The more important it is to protect the data and the more powerful your enemies are the better protection you should use. On the other hand, if you only want to protect against opportunistic casual prying by generally non-interested and less powerful enemies then it may be permissible to go for a lighter and less expensive protection scheme. This is analogous to other security/safety trade-offs that we do all the time. The main difference is of course that it is much harder to evaluate the risks and correspondingly to judge what constitutes an adequate protection level for data objects.

To enforce the data privacy property one can use different techniques. A data file on a hard drive is dependent on the operating system/file system being able to restrict access to the object. The operating system/file system can protect the data file in several ways depending on how strict the privacy requirement is, ranging from simple file system attribute settings to sophisticated use of data encryption techniques.

For objects in transit over a channel one should generally assume that the channel is insecure. That is, unless one takes explicit action to protect the channel against eavesdropping one should assume that the channel is open for all to read.

This also means that if data privacy is a requirement then one must explicitly protect the data when the data is in transit. The requirement applies to all 'open' communication channels, which in principle includes all radio channels and all IP-based channels.

1) *In the literature one tends to use the term entity to denote 'people or processes'.*

The protection in question is almost always achieved through use of data encryption techniques to provide the security service *Data Confidentiality*.

2.2 Identity- and Location Privacy

Identity- and location privacy are independent, but nevertheless related concepts. Again, the privacy property is about restricting access to information.

The *identity privacy* concept encompasses the right not to reveal you identity. *Identity privacy* is therefore concerned with restricting access such that only authorized parties will be permitted to know the identity. Of course, an entity may have several recognized identities and additionally there are other references and addresses that may also serve to identify the entity. The *identity privacy* issue is particularly acute for wireless communication or communication over open channels like the internet.

Location privacy is the right not to reveal the location of the entity. Note that the 'location' concept is not restricted to a physical location; it may well be a logical location.[2)] The *location privacy* property is not too relevant if the location never changes. For instance, a stationary object does not benefit nearly as much from location privacy as does a mobile subscriber. However, as is discussed in the article by Øverlier and Syverson on page 52 [2] even stationary entities (servers) may require a certain level of 'location' privacy.

The *location privacy* property is associated with *identity privacy* in the sense that to know the position of an identified entity may have much more value to an intruder than just to know that there is an unidentified entity at the same position. The article *Subscriber Privacy in Cellular Systems* on page 39 [3] takes a closer look at the current addressing model in cellular systems and the associated identity- and location privacy issues. The article also discusses new solutions to the privacy problems in the wireless context.

2.3 Movement- and Transaction Privacy

In the previous subsection we briefly introduced the identity- and location privacy concepts. It is possible to take these concepts further and add a temporal dimension, i.e. to establish a time series of observations.

Movement privacy is your right not to divulge your movements. For a cellular subscriber this means that no-one should be able to track your movements without your authorization. You may provide authorization (user consent) and the commercial so-called *location based services* (LBS) depends on you doing so. Movement privacy would also include data protection of your movement profile. This would include the property that GPS systems installed in cars do not leak information about the routes travelled.

It is noted that the movement privacy concept is not necessarily limited to identified entities. For instance, in military intelligence it may be of value to track the movements of entities even if their identities are unknown.

Transaction privacy is your right not to divulge your transactions. The concept is broad and could include any transaction that you would want to keep private. For instance, you may not want to divulge that you made a call to a certain person at a certain time. Or you may not want your spouse to know that you meet a certain person at a specific place/time. *Transaction privacy* is related to identity privacy and to location/movement privacy in the sense that if your movements are known then a lot of transaction information may be inferred or deduced. This is particularly true if one is able to correlate the information with information on the whereabouts of other entities or notable events etc.

2.4 Management of Privacy Sensitive Data

Management of privacy sensitive data is an important issue. Much of the data have limited lifetime and are subject to change. Information like your medical history is not static and information about your marital status may change and become outdated.

You obviously should have the right to ensure that the stored information is correct and you may have the right to ensure that the data is kept up to date. The authors of this article have surnames that frequently get misspelled. It is telling how difficult it can be to correct this information. The initial data entry may be a breeze, but corrections are often hard to achieve and there are often bureaucratic obstacles to be overcome.

Then there is the issue of ownership. Who owns the private data concerning you? The ownership issue concerns the right to copy and distribute the data and it concerns the right to erase the data.

For instance, if you decide to become the member of an organization then you may have to give up some personal data to the organization. Contact data (street address, phone number, email address etc) are almost certain to be included in the data you must submit. It is not uncommon that they ask you for the right to

2) *This could be expresses as a relative location with respect to other entities or as an address/reference in a database etc.*

sell the contact information to other companies/organizations. However, if you agree it is very hard for you to later ensure that the contact-data copies are kept updated and used only according to the agreement.

And what happens when you decide to quit your membership (or subscription or whatever). Under most jurisdictions the organization/company is allowed to keep the data for a certain period of time. Many organizations and companies will no doubt honour your privacy rights[3)], but many don't really have the proper data management in place and so you risk that the data is never erased. And, obviously, there are those organizations and companies that completely disregard your privacy rights.

A point in case with respect to management of privacy sensitive data is illustrated in the paper *Privacy and Unsolicited Commercial E-Mail* [4], where the authors found that although the great majority of websites that required registration information seemed to respect the restrictions it took only one rogue website to destroy your privacy. In the described experiment, the website in question distributed the contact data to spammers and so the effect was very visible and indeed very annoying (almost 500 spam emails were received during a five week period subsequent to the initial privacy violation). The example is a typical *tragedy of the commons*[4)] example, in which the conscious effort of the vast majority of the sites was completely wasted by one rogue site.

This aspect is very typical for privacy sensitive personal data. You are therefore well advised to be careful with whom you trust to know your personal data.

3 Real-World Privacy Threats

3.1 Your Mobile Phone and Other Personal Gadgets

3.1.1 The Mobile Phone

Is your mobile phone a threat to your privacy? The answer is that it clearly has the potential to be a threat to your personal privacy [5]. One aspect is that you may potentially be tracked. In fact, knowing where you are is one of the mail system-internal functions in a mobile system. This capability has been refined to provide location determination when making emergency calls. The capability is also used during lawful interception, but the most common usage will be (or become) the so-called *location based services (LBS)*. Abuse of LBS would be a threat to your identity- and location privacy. It is also, to a limited extent, possible to track users by observing the system signalling during execution of the access procedures.

For modern mobile systems like UMTS the threat against your data whilst in transit is minimal. The over-the-air protection is highly unlikely to be the weakest link (see [6] for an introduction to UMTS access security). The chance that your personal privacy is violated during data transfer is therefore diminishingly small[5)].

The real privacy threat to your data is when the data is stored in your mobile phone. Mobile phones these days have considerable storage capacity; not only do they store your contact lists etc, but they can also store large amounts of pictures and short video films. As the Norwegian actor Kåre Conradi got to experience first hand, it can be rather embarrassing to lose your mobile phone (the phone was stolen). Mr. Conradi had used the mobile phone camera to take a number of full frontal pictures of both himself and his girlfriend[6)]. The pictures were subsequently distributed on various internet sites. It is noted that publishing or otherwise spreading the pictures is considered illegal under Norwegian law. Another high profile mobile related privacy violation is reported to have happened to Paris Hilton[7)].

The modern smart phone is not only a mobile phone; it is also a full fledged computing platform. The phone can download and run applications. This makes the mobile phone susceptible to attacks that have previously been reserved for the personal computer. Some of these attacks are directly targeted at the user's pri-

3) *But what happens if they go bankrupt or are threatened by bankruptcy? Chances are that your data are sold off like any other asset.*

4) *"The tragedy of the commons is a type of social trap that involves a conflict over some resources between individual interests and the common good. The term derives originally from a parable published by William Forster Lloyd in his 1833 book on population. It was then popularized and extended by Garrett Hardin in his 1968 Science essay "The Tragedy of the Commons". However, the theory itself is as old as Aristotle, who said: "That which is common to the greatest number has the least care bestowed upon it."" Source: Wikipedia, http://en.wikipedia.org/wiki/Tragedy_of_the_commons.*

5) *If your GSM operator still uses the COMP128 authentication and key agreement functions (location in the SIM card) and/or if your operator still permits use of the broken A5/2 cipher algorithm (in the mobile station (MS) and base station transceiver (BTS)) then the data confidentiality of your GSM connection may be compromised. The attacks are for real, but you are still unlikely to ever experience an attack. For more information see [7]. Note also that in some countries encryption is prohibited. Then all bets are off.*

6) *An account, in Norwegian, is given in the net edition of the newspaper* ***VG*** *at http://www.vg.no/pub/vgart.hbs?artid=135588.*

7) *Reported amongst others at* ***The Register*** *at http://www.theregister.co.uk/2005/02/21/paris_hacked/.*

vacy. As already mentioned, the contact list is potentially a very private item. Then there is your call lists (time/date, duration and called-to number etc), the contents of your messages (SMS and MMS), all potentially very private and sensitive data. So the mobile phone, being a highly personal device, is likely to contain a lot of private information. This of course, has triggered the development of dedicated spyware targeted at these smart phones. A company called Retina-X Studios[8] has developed a Trojan spyware application called **Mobile Spy**[9]. This application will conceal itself after being installed and will report call activity and messages to a **Mobile Spy** account.

Quote (from the Retina-X Studios Mobile Spy homepage):

> *Mobile Spy is the next generation of smartphone spy software. Do you suspect that your child or employee is using your Windows Mobile phone inappropriately? If yes, then this software is ideal for you. Install this small program onto your compatible phone to begin recording.*
>
> *The software does NOT show up in the Windows Mobile task manager. It also does not rely on the phone's call and message logs to record activities. So even if the user tries to delete their tracks, the data will still be retained inside a small hidden file which will be uploaded to your account.*

Some of the claimed usage areas are *Enforce Employee Cell Phone Policy* and *Catch a Cheating Spouse*. While we have some sympathy with people being cheated upon, the installation of this application is nevertheless a gross invasion of a person's privacy and it is likely to be illegal in most jurisdictions. People with Symbian-based smart phones don't have to worry about **Mobile Spy**. They need instead to be worried about products like **FlexiSpy**, which if anything has even more potent spying capabilities[10].

The threat to your mobile phone doesn't stop there. The phones can also communicate using short-range communication, notably by IR, Bluetooth and RFID[11]. This can be quite handy for exchanging files and contact information etc, but as with most of the 'nice-to-have' functions there are pitfalls if used improperly. A number of high profile Bluetooth attacks have been demonstrated. The most infamous is perhaps the Flexilis **BlueSniper** demonstration, featuring a mean looking Bluetooth eavesdropping gun[12].

3.1.2 Other Personal Gadgets

People in the digital society seem to surround themselves with digital gadgets. An example is the new digital cameras with WLAN functionality and a GPS receiver. These cameras can tag your pictures with date/time and location and transfer the picture files wirelessly. One camera capable of doing this is the Ricoh Capelio 500 SE camera[13].

Other examples of course are the newer mobile phones. For instance, both the Sony Ericsson P990i phone and the Nokia N95 phone have these features, and they also feature MP3 players and considerable storage capacity.

There are also dedicated music player devices like the iPod or MP3 players. These devices, apart from playing music can also store any kind of file, including private pictures etc. Needless to say it may be quite devastating to lose such a device, not only from the loss of personal privacy but also from loss of irreplaceable data.

Newer devices now include biometric access control, password schemes etc to protect your data (see Figure 1 for an example). The security schemes, if implemented properly, will at least protect your data from falling into other people's hands. But not all products live up to the expectations. In a review of a product called the **Secustick** memory device the **Tweaknet** investigators found that the protection could easily be circumvented and that the claims made were false[14]. (It should be noted that the **Secustick** device has nothing to do with the device depicted in Figure 1).

---

8) *The Retina-X Studios homepage is located at http://www.retinaxstudios.com/.*

9) *The anti-virus company F-Secure categorizes **Mobile Spy** as a Trojan. Symantec, another anti-virus security company, is equally adamant in its damning of the **Mobile Spy** application. Both companies will add **Mobile Spy** detection to their software.*

10) *FlexiSpy is a product of a company called Vervata (http://www.vervata.com/) . The FlexiSpy homepage is at http://www.flexispy.com/.*

11) *See "Nokia brings RFID to mobile phones". http://www.vnunet.com/vnunet/news/ 2124563/nokia-brings-rfid-mobile-phones.*

12) *The story is featured on several homepages. One interesting example is the "How to: Building a BlueSniper Rifle …" articles at **smallnetbuilder.com**. The homepage for the first part is at http://www.smallnetbuilder.com/content/view/24256/98/.*

13) *http://www.ricoh.com/r_dc/caplio/500se/features/*

14) *See the story at http://tweakers.net/reviews/683.*

Figure 1 The Sandisk Cruzer Profile memory stick w/fingerprint identification

3.2 Threats against Your Personal Computer

3.2.1 An Insecure Operating System Cannot Protect You

It is well known that not all operating system (OS) are well secured. The insecurity stems from a long list of factors and the complexity of the systems certainly play an important role.

Windows is by far the most common OS found on the desktops/laptops. However, it is naïve to think that the Windows operating system family is the only one vulnerable. It is exceptionally difficult to construct a user-friendly feature-rich off-the-shelf operating system, and although Linux, MAC OS X and others seem to have done much better in terms of security it is still the case that these operating systems also have their share of security problems. Windows is however the OS that most users have installed on their laptop/desktop and so it is a tempting target.

If the operating system integrity is broken then your system may be compromised. Your system is then vulnerable to attacks. So vulnerable in fact, that on an average you may not have time to run the security upgrades before the system has been successfully attacked[15]. Many users aren't even aware that they must regularly download and install security upgrades and patches.

Once the system is compromised there is a lot that can be deduced about you and your behaviour. Modern operating systems feature **System restore** functions (with system configuration data), **hibernation** and **swapfiles** (these contain copies of system main memory), **history** functions and **search indexes** etc. All these functions contain traces of your activity on the system and a skillful attacker will be able to exploit these data sources. Of course, the skilled attacker is also able to exploit the fact that **deleted** files on your hard drive will normally not be deleted; most of the time the operating system merely deletes the reference to the file and not the file itself. System object databases, like the Windows **Registry** is also a rich source of potentially private information.

The attacks to exploit the above data sources are not easy to execute and the attacks are in practice only feasible when the attacker has gained physical access to your computer. This is nevertheless a real threat since an alarming number of laptops are lost and stolen each year[16]. The use of hard drive encryption is helpful to mitigate the effects, and such software is expected to become prevalent on business laptops[17]. The methods used by malicious hackers are not exclusively used by the bad guys; the same techniques are also used by police authorities in forensic investigations of computer evidence (see [8] for more background).

3.2.2 Office Applications and the Perils of Meta-Data

The office applications, unbeknown to most people, store a lot of so-called meta-information in the office files. For instance, MS Word files can store a lot of information about older editions of the file. According to Microsoft [9] Word documents can contain a lot of metadata, including

- Hidden Revision Logs
- Comments, Keywords, Subjects, and other properties
- Recent Hyperlinks
- Last Saved Date, Last Printed Date
- Last Edited By Information, Total Editing Time
- Revision Count.

The information seems innocent enough, but when the document is distributed on the Internet you may not want all this information published along your document. Microsoft has published a Knowledge Base article on the subject [9] (applies to Word

15) *In 2004 SANS Institute Storm Center reported a survival time of 20 minutes for an out-of-the-box Windows XP systems. The numbers have now improved, mostly due to the firewall included with newer Windows systems. SANS ISC maintains a* **survival time** *homepage at http://isc.sans.org/survivaltime.html. The survival time page is updated regularly.*

16) *See the story "Ernst & Young laptop loss exposes 243,000 Hotels.com customers" at http://www.theregister.com/2006/06/01/ey_hotels_laptop/ for an illustrating example.*

17) *Encryption directly on the hard drive is preferable. Hard drives with these capabilities are now about to enter the market. One recent example is the Hitachi Travelstar 7K200 encrypted hard drive.*

2003). The problems are not limited to MS Word. According to a Gartner article, *Microsoft Office metadata: What you don't see can hurt you* published on the **TechRepublic** homepage[18] the problem also exists for other Office application file formats, notably Excel and PowerPoint. The problem is not isolated to the MS Office file formats; it does for instance also affect the Adobe Acrobat PDF file format.

Even if we have portrayed metadata as a privacy curse in the above section, metadata is of course also very useful. Many advanced features rely on the presence of metadata and in many circumstances the metadata is required to be present. So, the solution isn't to get rid of metadata as such, but rather to control access to it. This is particularly relevant for the finished document which is ready to be published.

Are there any solutions to the metadata problem? Well, for Word files you may follow the advice in the MS Knowledge Base articles and do the metadata-cleaning job manually. Then there are the helper applications. Examples here include the free **Doc Scrubber**[19] application (which, unfortunately, does not seem to be maintained anymore) and commercially available solutions, like the **Metadata Assistant**[20].

Metadata is an essential and integral part of the new Web 2.0 vision. This may provide us with smarter and more customized applications and services, but the price to pay may very well be measured in terms of lost personal privacy. The Web 2.0 vision is centred around the web browser as the one-and-only client. This then brings us over to the subject of your browser vs. your privacy.

3.2.3 You, Your Browser and Your Browsing Habits

The following questions are related to your privacy: How much does your browser know about you, how much does it expose you through your browsing, and how much can you trust your browser?

The lure of **auto-complete**, **history** lists and **remember password** functions etc is considerable. For a personal computer user those functions are indeed handy and can make your browsing a much more pleasant experience. However if you lose your computer or it is hijacked by malware or simply accessed by other people, then all of your browsing habits are exposed. There is of course a remedy to this. In addition to turning those convenient functions off you can also clear the private data. In Firefox v2 you can for instance execute the **Clear Private Data** function (**Tools** menu-bar or *ctrl-shift-del*) to get access to a menu where you can selectively delete the private data. Similar functionality can also be found in Internet Explorer, but not quite as accessible as in Firefox (see **Tools** in the menu-bar in IE 7; in IE 6 the functionality is harder to find and is spread across several menus).

The Firefox web browser has been a most welcome competitor in the web browser area. It has presented many new innovative features and has generally been more secure than MS Internet Explorer. The Firefox browser was developed through a community effort and the source code is the public domain. That is, anyone who cares to can download the browser code, make a little change and recompile the browser into your personal edition. This is nice and very convenient. However, it also highlights that there is a serious trust issue here. Can you trust your software?

The old UNIX guru Ken Thompson wrote a very interesting paper entitled *Reflections on Trusting Trust* [10] which he used as his inauguration speech paper when he received the ACM Turing Award in 1983. In this paper he demonstrated how he could, in a convoluted and covert way, recompile the UNIX **login** program such that he would be able to log on as super user. The feat was achieved in such a way that the **login** source code left on the system would be the original code and thus even source code inspection wouldn't reveal that **login** had become a Trojan program (recompiling login wouldn't help either!). Ken Thompson's assertion was that you should not trust the software unless you created it yourself[21].

In practice we must be a little more pragmatic and we therefore generally trust our software suppliers to at least try their best and we trust them not to wilfully try to deceive us. But then what happens with open software? We may feel that we can trust the team that originally developed Firefox (or those behind our Linux edition for that matter), but does that mean that

---

18) *Online version available at http://articles.techrepublic.com.com/5100-1035_11-5034376.html. To download the full article you must register (free, but as always there is a privacy price to be paid). TechRepulic also contains a whitepaper on the subject called "Meta Data Can Harm Your Business" (http://whitepapers.techrepublic.com.com/whitepaper.aspx?&q=meta+data&docid =167360), but you need to register before you can download the article.*

19) *The DocScubber homepage is at http://www.docscrubber.com/.*

20) *Metadata Assistant is a product of PayneGroup. More information at http://www.payneconsulting.com/products/metadataretail/.*

21) *The actual statement was: "The moral is obvious. You can't trust code that you did not totally create yourself." [10]*

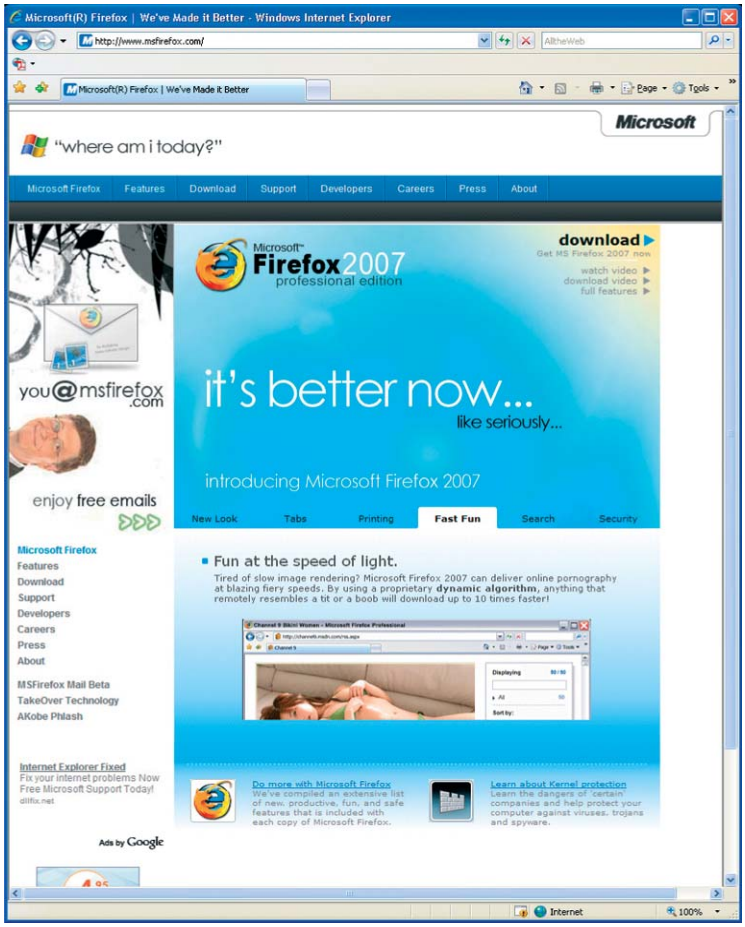


Figure 2 MS Firefox 2007 (http://www.msfirefox.com/) Certainly not from Microsoft. But, who provides your browser? Can you trust them? With Firefox code being available any competent hacker can put together his/her own Firefox distro (see http://developer.mozilla.org/en/docs/Build_Documentation for documentation on how to do this). The real Firefox is available from multiple sources and servers, amongst them http://www.mozilla.com/en-US/firefox/. The default is for an American-English version, but a large set of other languages are available

we can trust the software we actually downloaded. How do we know that we got the original edition and not a tampered with copy? Figure 2 provides an interesting take on this issue. So, the question remains in full force: Can you trust your software provider? And is open source the best option seen from a security/privacy perspective?

The open source community often argue that open source software is more reliable and that it can be trusted since it's readily available for all to inspect. In the paper [11], Ross Anderson compares the open and closed source development processes and shows rigorously that each process has its own pros and cons, and there is no evidence that open source produces software with fewer security holes and therefore could be called more trustful in an ideal world. In his words "… whether open or closed systems are more secure in a given situation will depend on whether, and how, the application deviates from the standard assumptions".

3.3 Why Should You Trust Them

3.3.1 Digital Privacy and Self Illusions

Vanity and futility reigns, or so it seems. The need to be *seen* and *recognized* is a deep psychological need and in a digital society that need can be satisfied by digital means. So phenomena like *Facebook*, *YouTube*, *MySpace* and similar do serve a need and can probably be explained correspondingly.

However, seen from a privacy perspective there are big and deeply disturbing aspects of the digital self portraits and self presentations that people put out on these websites. The problem is firmly rooted in the data management problems outlined in the subsection *Management of Privacy Sensitive Data*. Once you have published the data you essentially lose control of your personal data. This can have severe consequences to the individual. Some of these sites add material and update your personal data without informing you. All aspects of your web presence may end up in a *Facebook* page, with or without you agreeing to it. This would include information that you'd rather have people forget about, but there is no forgetting of embarrassing moments on the internet.

People have started to experience the negative sides to full publicity by now. It is not uncommon for employers to conduct an internet scan to see what pops up when searching for your name. Those party pictures that were so funny then are now a liability to you, and you may well find that you didn't get the job due to a silly video you posted on *YouTube*.

The problem is not limited to *Facebook*, *YouTube*, *MySpace* and the likes. Many people run personal web pages with a lot of private pictures etc, many schools publish pictures of the pupils (some have rules not to identify the pupils, but can we really trust the average public school and/or teacher to be experts on data privacy?) at their websites and many employers publish contact information of their employees (in addition to being downright silly with respect to hostile recruiting it may also violate the personal privacy of the employer).

3.3.2 Search Engines

Web services like *Facebook*, *YouTube* and *MySpace* may or may not honor your personal privacy the way you like. However, even if we assume that you would have full administrative control over everything they publish about you it still does not mean that you can revoke and deleted information you have published.

The search engines will see to it that the revoked/deleted information is still present. Sometimes this serves an important purpose and it may indeed be useful to be able to retrieve 'archive' information. The search engines then serve a digital library like services. Still, there clearly is a personal privacy aspect to this and as of writing this, there is a discussion on how the search engines may protect your privacy. However, while the search engine companies may be willing to discuss the topic they clearly have a different perspective from you and me. In the recent online article *Google plays cat and mouse with regulators* [12] the issues are highlighted, and it should be clear that unless the regulatory authorities are strict we all face search engine related privacy problems in the future.

3.3.3 Want Some Cookies?

A cookie is a small text file. It is used by web servers as an easy means to add state and context to a web surfing session. The cookie is issued by the web server and is stored locally at your computer. The cookie is returned to the server by the browser upon subsequent visits to the server site.

The cookies are used for authenticating, tracking, and maintaining state information about user activities. The state information may include user identity, time of last visit, site preferences, the contents of electronic shopping carts etc. The lifetime of cookies can be set, but the expired cookies may remain on your computer. The cookies can also be allowed to work over a set of web servers.

Cookies are convenient and useful, and can add significantly to a positive user experience. However, by their very versatility and utility the cookies can also by used to track your browsing habits. In fact, many web pages use the cookies not so much to aid in navigating their site as in attempting to sell you goods or to redirect you to other sites (which may pay for the service).

This may not at all be in your best interest. The cookies are then effectively turned into *adware*[22)] tools.

You can configure your browser to disallow cookies or to impose restrictions on how the cookies are used. However, given that so many legitimate sites use cookies dynamically and extensively it can be a never-ending exercise to configure the browser to protect you. It is then probably a better solution to regularly use one of the available anti adware/spyware tools. A particularly popular tool against adware is the *Ad-Aware* application[23)]. Another tool worth using on a regular basis is the *Spybot Search and Destroy* program[24)].

It should by noted that many of the peer-to-peer file sharing tools are notorious for distributing spyware/adware and that furthermore many fraudsters use these tools for tricking users to download adware/spyware and other malware.

3.3.4 Gone Phishing

Phishing is an activity where a fraudster tries to acquire sensitive/private information. They commonly use social engineering techniques. The most sought after information is usernames, passwords and credit card details etc. A well know phishing technique is to masquerade as a trustworthy website. This includes internet/online banks, auction companies like eBay and other trustworthy sites where you may be compelled to leave sensitive data. Phishing is typically carried out using email and the users are conned to login or otherwise convey information at a website.

A number of the scams involve distributing false emails with an announcement from an internet bank or credit company that there has been a security problem. The emails may look quite professional (but not always) and they then (conveniently) provide you with a link to a login page. The login page is of course a fake, but it can look very similar to the original site that they would have you believe that you log into.

Our best advice is: **Do not reply or act on such requests!**

To make the web addresses etc. look genuine they often use addresses that by a superficial glance look the same as the original. An example of address faking:

A) http://www.telenor.com/telektronikk/
B) http://www.telenor.net/telektronikk/

---

22) Adware *is software which displays or downloads advertising material to your computer without your consent. It is typically difficult to get rid of the* adware *and it shares many traits with its more sinister cousin* spyware*. The term* spyware *is used for software that collects personal information about users without their informed consent. The spyware often uses stealth techniques to hide its activity or pose as a legitimate application. Note that spyware may be used for* adware *purposes, but that* spyware *may also be used for* phishing*, identity theft etc.*

23) *The Ad-Aware tool is made by Lavasoft AB (Sweden). It is available for free personal use. More information at http://www.lavasoft.com/products/ad_aware_free.php*

24) *Spybot S&D is made by Safer Networking Limited and is available for free personal use. More information at http://www.safer-networking.org/en/download/index.html*

Address *B)*, which seems fine, is not a valid *Telektronikk* address. But, a fraudster could apply for this address and set up a fake *Telektronikk* homepage. Then they could harvest contact information (the *Subscribe* menu choice) and if they were clever they would even pass along the information to the true *Telektronikk* homepage. You may check who has registered a domain name by inquiring a **whois** server[25)].

The above example is somewhat contrived, but the problem is serious. The modern browsers are partly aware of the problem and they contain anti-phishing mechanisms. For instance, Microsoft Internet Explorer 7 contains a phishing filter (you can configure it in the **Tools** menu-bar). Anti-virus and anti-spyware products also regularly feature real-time protection against well-known phishing scams.

3.4 Using WLAN, Are You?

3.4.1 Is Security Turned on?

In the last few years an enormous number of IEEE 802.11 WLAN access points have been installed, both in office environments and in people's homes (commonly as an extension to ADSL network access).

Both from a security and privacy point of view there are a number of concerns with the use of WLAN. In small offices etc and in people's homes these installations are mostly unmanaged. That is, they are almost always installed with default configuration. The default configuration is quite often not a secure configuration. That is, it is common to leave out authentication and encryption in the default configuration. This is understandable since it makes sense to check whether the basic installation works before configuring the access security. But, the problem with this is that many installations are left unprotected. All private data, often even including password, is then transmitted in cleartext. Anybody within radio access reading distance can then freely eavesdrop on your data. This means that your neighbours and any visitor to your neighborhood *may* see your data. Depending on configuration, this may even mean they will get access to your computer and your harddisk.

3.4.2 On the Road

Your WLAN card is normally configured to look out for WLAN access points. This feature can normally be turned off, but many never bother to do so. The WLAN card can also be set up to automatically connect with accessible networks. This is often the default setting.

It can be very convenient, but it can also be illegal, unethical and potentially quite dangerous. The convenient part is the easiest. You get access to the internet! The illegal and unethical part is based on the fact that you may access and use other people's internet connection without permission. In many countries this is illegal and it is still trespassing even if it may not be explicitly illegal.

The dangerous part is that someone may try to trap you. This can be done through dedicated attacks, but more likely the intruder is opportunistic. They put out the WLAN bait and inspect what they hook. If they find that your operating system or browser is not updated with the latest security patches they may easily gain full access to your computer. No hope of personal privacy after that.

3.4.3 Broken Security

It is hard to assess the actual risks we run with respect to eavesdropping. This is of course part of the problem, since we're not generally inclined to do much in terms of protection unless we understand the problem.

This goes for WLAN security as well. The original cryptographic WLAN protection scheme, called Wired Equivalent Privacy (WEP), is broken and has been known to be so for a number of years. Still, the fact that a cryptographic scheme can be broken is in itself not the same as it being possible to mount practical attacks. Attacks must be cheap and easy to mount before really becoming significant to most of us. Attacks that scale are also something we need to worry about.

When it comes to WLAN the attacks against WEP do not scale terribly well as they are physically restricted. The earliest attacks took several hours and required a lot of traffic on the link. However, the most recent attacks are very fast, and a successful

---

25) *Of course,* ***whois.com*** *is taken (by what looks mostly like an advertising outfit!). The Icelandic Network Information Centre (NIC) operates a* whois *server at* ***who.is*** *(http://www.who.is/). The* ***.com*** whois *provider is Network Solutions (http://www.networksolutions.com/whois/index.jsp). Likewise there is a* ***.net*** whois *service at http://www.whois.net/index.php. It is noted that the different domain name authorities provide* whois *information in several incompatible formats, and that you may have to go to a local* whois *authority to get the information. The* whois *servers may or may not relay information from other* whois *authorities. This is partly a conscious decision as it does make it harder for the spammers to collect domain names.*
By the way, the telenor.net *domain is taken by Telenor.*

attack could be conducted literally within minutes[26]. Then the attacker will have got hold of the secret key and every bit of data can effortlessly be deciphered (note that this also holds for previously recorded data under the same cipher key; thus even previously transmitted data may be at risk). The paper *Breaking 104 bit WEP in less than 60 seconds* [13] provides the technical background information on the latest attacks.

The concerned reader should know that WPA and WPA2[27], which are the replacement algorithms for WEP, are believed to be secure for now (though we advise use of WPA2 instead of WPA).

3.5 Digital Rights and Digital Wrongs

With digital content (music, pictures, films etc) comes the issues of digital copyrights. The fact that the authors or content owners have intellectual property rights to the material is not generally disputed, although there obviously also are people that dispute this right altogether.

The problem with digital rights is of course how to enforce the rights. The problem stems from the fact that it is essentially without cost to copy and distribute the fully digitalized material. This makes it very easy for people to make perfect copies and distribute the copies to their friends. The problem is exacerbated by the fact that making a limited number of copies may be perfectly legal. The copies may be permitted for use on different playing devices (by the same person) and for having a backup copy (in case of disk crash etc). It is not easy to enforce digital copy rights without either needlessly restricting the usage of the digital object or by being too intrusive in the enforcement.

When the DVD format was conceived the system was developed with a scheme called the Content Scrambling System (CSS). This protection scheme was intended to ensure that DVDs could only be played on DVD players with the CSS system. The story of *DVD Jon* and how Jon Lech Johansen and others were able to crack CSS is history by now, but it nicely illustrates the problems of digital content control. The story also illustrates the inherent conflict between customer rights and digital rights. For Jon Lech Johansen only developed a Linux driver to be able to watch his DVDs on his Linux machine. He did purportedly not steal any content, nor did he stand to gain commercially from the exploit (the code snippet was called DeCSS). He only wanted to watch content that he had already paid for. The story evolved into a legal case and Johansen had to fight his case in court (he was eventually acquitted). The Electronic Frontier Foundation (EFF) maintains a digital archive of the case at http://www.eff.org/IP/Video/.

Another more recent incident concerns the misconduct of Sony BMG in what is often termed the Sony DRM[28] Rootkit scandal. Sony had started to protect some music CDs with anti-copy software. There were two different schemes called **Extended Copy Protection (XCP)** and **MediaMax CD-3**, respectively. The worst scandal concerns the XCP software and took place during autumn 2005. The software installed itself when the CD was put in the CD player on Windows machines. The way the installation was done was decidedly covert and the software went to great lengths to hide its presence. It also interfered with the machine's capability for playing CDs. In particular, the software did not allow copies to be made of the DRM protected CDs.

The way the software hid itself was reminiscent of the so-called *rootkits*[29]. Indeed, the Sony DRM system was considered as spyware/malware by most of the anti-virus industry, including F-Secure and Symantec. To add insult to already sustained injury, Microsoft classified the Sony DRM rootkit as malicious when they reported that they would include XCP removal in their *Malicious Software Removal Tool*[30].

The story of the XCP software is fascinating and a useful account can be found at [14], and although this particular story ended with full retraction of the offending software, the problem largely remains.

So what has all this to do with personal privacy, you may ask? Well, many of the DRM systems are costly to develop and they tend to rely on online registration verification. Having developed a backdoor system that phones home regularly it is *very tempting* to include a little more than the copyright verification.

---

26) *The claims made are perhaps a little optimistic and the actual breaking time does depend on the amount of data transmitted over the link.*

27) *WPA and WPA2 are commercial brand names (Wi-Fi Protected Access). Both methods are derived from the IEEE 802.11i specification.*

28) *DRM is a common abbreviation for Digital Rights Management.*

29) *A rootkit is a software tool that conceals itself and other running processes, files or system data from the operating system. The rootkits will usually modify parts of the operating system, install itself as a device driver or pretend to be part of the operating system. Rootkits are generally considered to be so-called malware, but may in theory also be benign.*

30) *The Malicious Software Removal Tool is available for free at http://www.microsoft.com/security/malwareremove/default.mspx.*

Many of the schemes also require quite a lot of additional information to work, like the licence number of your operating system, the processor- and motherboard identity of your PC[31], the hard drive type the software is stored on, etc. Incidents have been reported where DRM schemes have reported a lot more back home than they really need to, and this is certainly a threat to your privacy. It could be relatively innocent like the 2006 controversy regarding the Microsoft Windows Genuine Advantage (WGA) phone home functionality. The WGA software, which was intended to combat illegal copying of the Windows operating system, called home to the Microsoft servers every day. The problem here is that the legitimate paying customer was adversely affected by this software and furthermore that Microsoft didn't tell you that the software would be phoning home. And how would we, the law abiding citizens, really know if the WGA tool does not spy on us? It is noted that offline based DRM solutions, for example solutions based on tamper-proof devices like smart cards [15, 16] would improve user privacy.

3.6 Pervasive and Invasive – The RFID Revolution

We shall not examine the topic of RFID and personal privacy here, but instead refer to the article *RFID and Privacy* on page 77 [17].

3.7 Data Mining

Data mining is the process of searching large volumes of data for patterns using various tools to categorize and correlate data. Thus, data mining is the process of transforming apparently unrelated data from large data sets into usable information. Data mining has also been defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data" [18].

This edition of *Telektronikk* features a separate article on personal privacy and data mining [19]. The article focuses on privacy preserving techniques that can be used to provide credible user privacy while still achieving the major goals of the data mining effort. For instance, while personal data may need to be taken into account during the data mining effort, there is generally less reason for the derived results to contain privacy sensitive data. Under these circumstances one may well be able to extract the desired information without violating the user's personal data privacy rights.

3.8 The Spy in the Sky

One area that could easily provoke a mild case of paranoia is the spy-in-the-sky scenario. We're probably all aware of the spy satellites operated by military and intelligence agency organizations. We're also aware that these satellites are very capable and able to capture a great deal of detail at ground level.

The emergence of services like Google Earth[32] has merely highlighted the issue, as commercial satellite image capturing has been available for some time now[33]. Now, the commercial operators do not off-the-shelf provide the same high definition pictures as is available to the military. Neither do they normally provide 'surveillance' services. However, the resolution can be quite high. One example is this quote by TerraServer on their homepage[34]: *"For our high-resolution imagery which is a mixture of aerial and satellite shots, this varies anywhere between 0.1 meters and 2 meters depending on the capabilities of the camera that was used to take the pictures."*

For political and legal reasons you may not be allowed to buy pictures from all locations. The commercial operators may respect restrictions imposed by the national authorities, particularly in the area of 'sensitive areas'. These sensitive areas would likely be areas of military interest, and while you may get hold of pictures of the area they may either be explicitly blurred or old versions[35]. Still, the military is concerned with this issue and the US National Geospatial-Intelligence Agency (NGA) (http://www.nga.mil/) has openly called for more control over commercial satellite picture providers[36].

So, should we worry about the spy-in-the-sky? Well, the issue is real but unless you have a paranoid disposition the personal privacy problems caused by the off-the-shelf 15-meter resolution pictures are really only a minor issue. The potential is there, but it

---

31) *The subject of having a unique serial number for each CPU is quite controversial and it seems that the CPU vendors no longer support this.*

32) *The Google Earth homepage is at http://earth.google.com/. Google Earth itself if a free service, but they also have paid companion services (higher resolution, more tools etc).*

33) *EarthSat for instance, has been around since 1969. Homepage at http://www.earthsat.com/.*

34) *The main TerraServer homepage is at http://www.terraserver.com/. The quote is from the http://www.terraserver.com/products/images.asp homepage.*

35) *Google Earth removed British military bases from maps of Iraq after allegations that terrorist used the maps for planning attacks. Source: The Register, http://www.theregister.com/2007/01/17/google_erases_brit_bases/.*

36) *The NGA concern is reported at Komo-TV site at http://www.komotv.com/news/tech/7404181.html.*

seems likely that the military and other powers-that-be will not allow high-resolution pictures to be sold off-the-shelf.

3.9 Guilty Till Proven Guilty

Ever got a speeding ticket from a speed control camera? If you have, then you know just how irritating that can be. However, from a personal privacy perspective there is no real issue. You break the law and then you have your picture taken. If you drive responsibly then no picture is taken. So, that's all fine then.

In Norway, and most likely other countries, the traffic safety authorities have carried out a pilot project measuring the average speed between speed control cameras. Technically, this is becoming a relatively easy exercise. One needs high-quality pictures and accompanying software to identify the car (colour, size, type etc) and the licence plate. The rest is just a matter of processing.

From a personal privacy point of view this approach to traffic safety is utterly wrong. The problem is that they take your picture even when you are driving within the speed limit in the event that you might break the speed limit (on average) later on. So, the speed control cameras will then need to take a picture of every passing car. You are then truly guilty until proven guilty.

In this particular case personal privacy seems to have won. The Norwegian Datatilsynet[37] (The Data Inspectorate), which gave permission to the pilot project, was in the end less enthusiastic about the project. So, it seems we will not see the proposed sign, shown in Figure 3, on Norwegian roads anytime soon.

Figure 3 The proposed sign for the average speed automatic speed control cameras

4 User Consent is no Silver Bullet

User consent is often seen as an acceptable approach to gathering and using privacy sensitive data. For the data collector's point of view *user consent* is an effective solution. However, it is questionable how effective the *user consent* option is with respect to protecting people and whether users are able to understand what they are agreeing to [20].

User consent comes in different forms and flavors. One very basic characteristic of user consent that must be in place is that there should be *informed consent*. That is, the average user must be able to understand what they are agreeing to. This means that the often highly technical or legalese texts that often are presented as 'information' to the users is missing the point. This problem is of course also exacerbated by the fact that a relatively large part of the population, including kids, will lack the literacy skill and the mental maturity to understand the ramifications of their decisions. So, while informed consent is a prerequisite for the acceptance of user consent, it is a very hard target to meet in many circumstances.

The acceptance criteria for user consent should also normally be based on so-called *positive user consent*. That is, the default action should be that the user does not give up any privacy rights. Only when the user takes unambiguous action is permission given. However, in the real world this can be a bit restrictive and it is common to allow the default choice to include acceptance of use of private data.

The *implied/associated consent* can also have practical benefits to the user. They will then not have to be bothered with additional 'please provide your consent' dialogues. And, for associated services or for services within the same domain that may be considered a benefit for the user provided the associated consent is within the intentions and scope of the original consent.

One must be realistic here. Bothering the user with a large set of consent questions does not really serve any purpose. It simply wears down the respect for the consent question, and the users then tend to answer **YES** almost on auto-pilot[38].

---

[37] *http://www.datatilsynet.no/.*

[38] *The authors have a suspicion that this is learned behavior from the personal computer environment. During installation of programs or features the users are commonly asked a whole series of "Will you permit/accept this-or-that: YES/NO" questions. From experience the users then tend to learn that answering NO means that the program/service will not install. Thus, one learns to disregard the question and simply answer YES to any such question.*

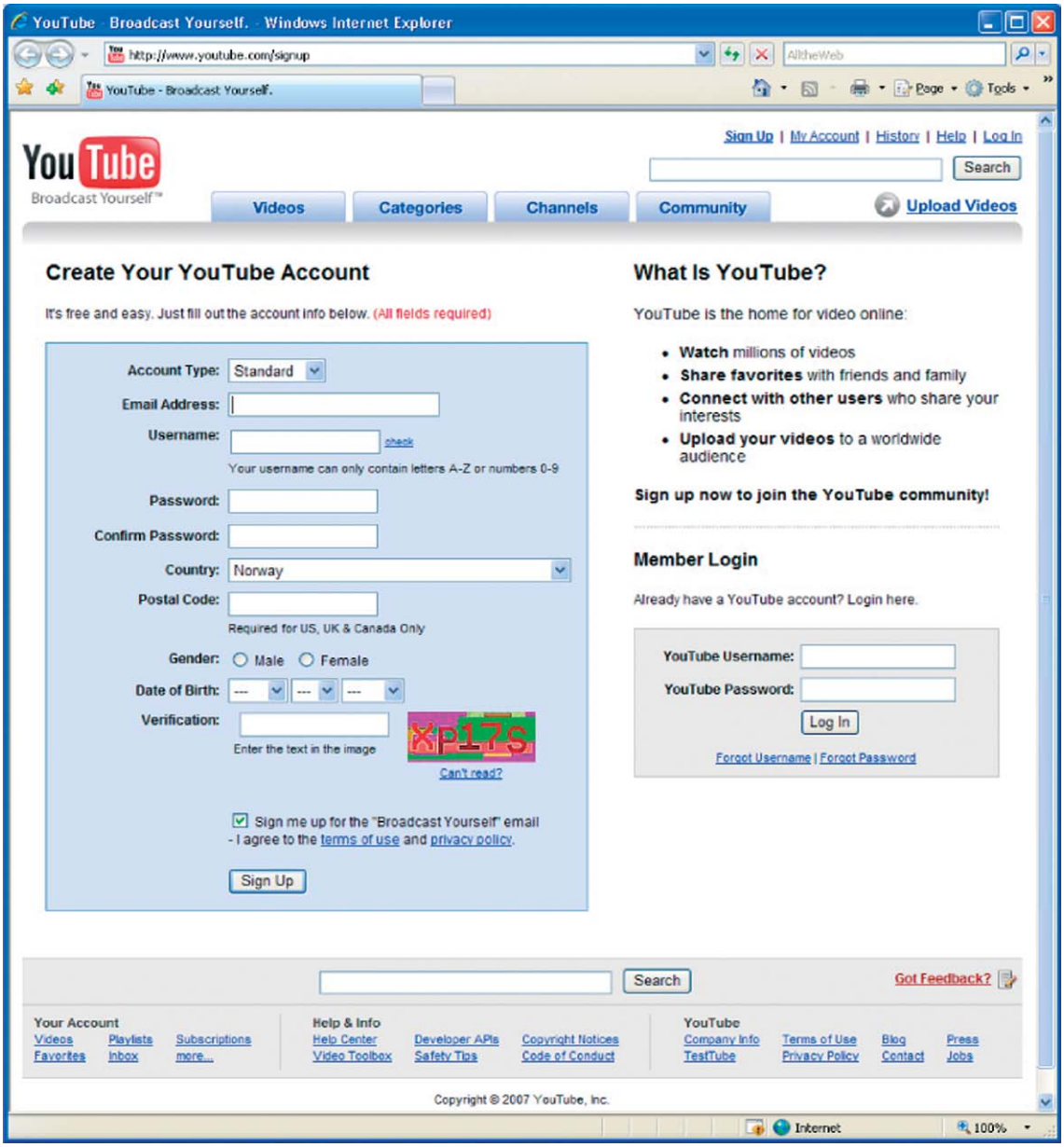


Figure 4 YouTube account registration (http://www.youtube.com/signup). The Sign-me-up for the "Broadcast Yourself" email was ticked off as default. YouTube requires that you enter date of birth and gender

Another problem with user consent mentioned above is that many users seem not to understand what they are agreeing to. Furthermore, most users seem quite willing to give away rather sensitive data for flattery (or recognition) or some other minor benefit.

For instance, when **Infosecurity Europe** in London conducted a survey on users' willingness to reveal their passwords almost 2/3 gave away their password for a "chocolate and a smile"[39]. Now, most people realize that the password is **not** something you should reveal, yet a stunning 64 % actually did just that (assuming they were telling the truth, that is). Many web sites provide other incentives to the user to have their personal data; the most common being free access to the web pages or to software etc. Figures 4, 5 and 6 give a few examples of the amount of personal data some sites require, ranging from a lot of detail to nothing, as in Figure 6 (though you may submit contact information if you want to be contacted). The amount of data that some sites want to register is surprising, particularly when considering how poor the data quality assurance generally is[40]. Many sites lack even the most rudimentary sanity check on the

39) *The press release from Infosecurity Europe is available at http://www.infosec.co.uk/page.cfm/Action=Press/PressID=640.*

40) *It is very easy to register erroneous data. However, to correct the data later on is often much more difficult, let alone of you want to discontinue the service and want to erase the data.*

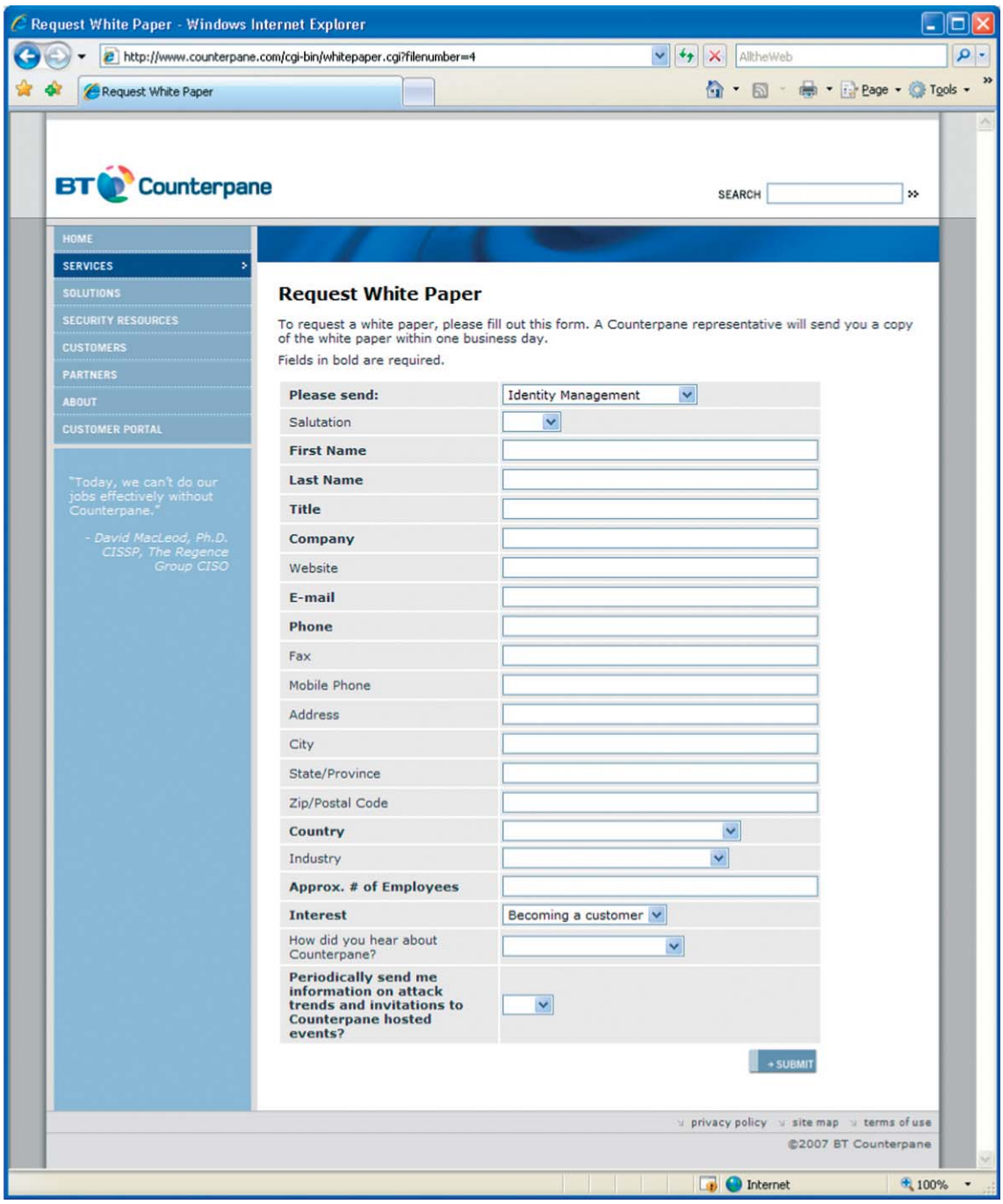


Figure 5 BT Counterpane registration for access to a whitepaper. Ironically the access is to a whitepaper named "Data Privacy & Protection". The BT Counterpane homepage is hosted at http://www.counterpane.com/index.html

input or does not accept perfectly valid input (usually because the site is US centric and unaware of all valid formats).

5 Where Personal Privacy Ends – Legal Aspects

Personal privacy is important, but it does not take precedence over state security. Therefore there exist laws and regulations that govern when and how the national state is permitted to violate your personal privacy. This is captured in what is known as *Lawful Interception (LI)*. For more information on lawful interception the reader is directed to the article *Lawful Interception* [21] on page 33. The LI measures are targeted and only applied when there is reasonable cause for suspicion. One also has the so-called data retention type of surveillance. The data retention approach is focused on data mining of call data (not the content per se, but where, when and to/from type of data), but is applied indiscriminately and your call data are collected even when there is no suspicion against you. In the article *The EU Directive on Data Retention – An End to Justify the Means* [22] on page

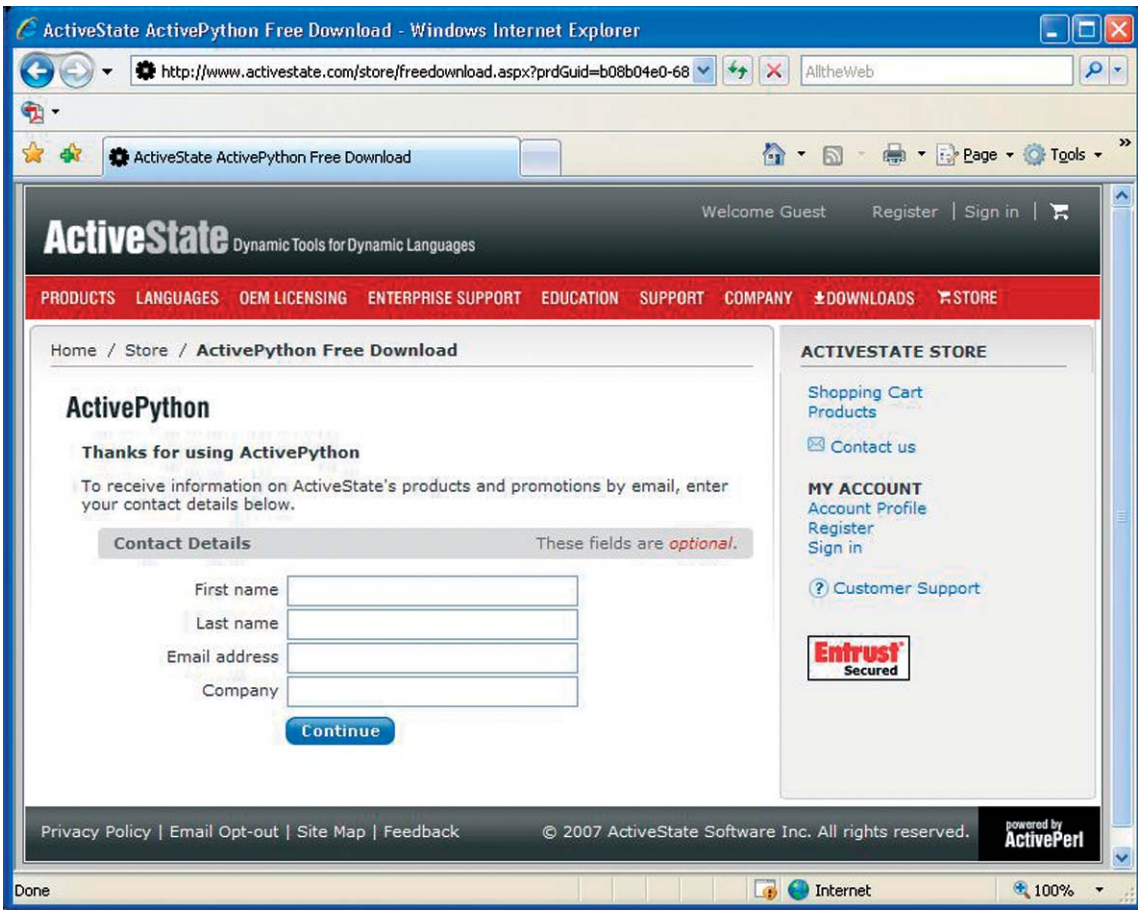


Figure 6 User Consent at ActiveState (www.activestate.com). You need only provide registration data if you want to receive information from ActiveState. To get the free software without registering the user would just click on ***Continue*** *without filling in the* ***Contact Details*** *form*

31 the author takes a critical look at the EU Data Retention Directive.

In your home country you may trust the authorities to respect your personal privacy. But what happens when you move abroad? Then you will certainly be under a foreign jurisdiction. This, literally, goes with the territory. So far so good, but what happens when some virtual service is located in another jurisdiction? And how will you know where the service is hosted? Where are the Google servers located? Where are the Facebook servers located? Your privacy may therefore by subject to rules and regulations that you know nothing about. Harmonizing of privacy protection rules would have been nice, but there is a considerable difference of opinion about this subject and people's tolerance for privacy violations vary considerable.

6 Summary

This article has attempted to give an overview of some real-world personal privacy problem areas. There are many challenges, and new technology can be our biggest problem but ultimately also the only viable path to a reasonable and well-balanced solution. Many initiatives are being taken to solve the various personal privacy problems, and for several of the problem areas one now finds that the privacy enhancing technologies are reaching a relatively mature status. Still, much more work is needed in this area if one wants to address the problems in a reasonably uniform manner. Many large companies have appointed a Privacy Ombudsman, which then is responsible for protecting the privacy of the customers and clients (see the article *Privacy and Protection of Personal Data* [23] on page 27 by the Telenor Privacy Ombudsman for an example).

It is reassuring that various regulatory authorities now are beginning to press for development of more effective privacy enhancement technologies. A point in case is the recent initiative from the EU Commission [24] to encourage the use of privacy enhancing technologies to provide credible data privacy.

References

1 Tavani, H T, Moor, J H. Privacy protection, control of information, and privacy-enhancing technologies. *SIGCAS Comp. Soc.*, 31, 1, 6-11, 2001.

2 Øverlier, L, Syverson, P. Location Hidden Servers and Valet Nodes. *Telektronikk*, 103 (2), 52-60, 2007. (This issue)

3 Køien, G M. Subscriber Privacy in Cellular Systems. *Telektronikk*, 103 (2), 39-51, 2007. (This issue)

4 Jacobsson, A, Carlson, B. Privacy and Unsolicited Commercial E-Mail. In: *Proceedings of NORDSEC 2003*, Gjøvik, Norway, 15-17 Oct 2003, 1-12.

5 Gratzer, V, Naccache, D. Cryptography, Law Enforcement, and Mobile Communications. *IEEE Security & Privacy,* 4 (6), 67-70, Nov/Dec 2006.

6 Køien, G M. An introduction to access security in UMTS. *IEEE Wireless Communications magazine*, 11 (1), 8-18, 2004.

7 Barkan, E, Biham, E, Keller, N. Instant Ciphertext-Only Cryptanalysis of GSM Encrypted Communication. In: *Advances in Cryptology – CRYPTO 2003*, 600-616, Springer, 2003. (LNCS 2729)

8 Allen, W H. Computer Forensics. *IEEE Security & Privacy magazine*, 3 (4), 59-62, 2005.

9 Microsoft Knowledge Base. *How to minimize metadata in Word 2003*. Article 825576, Revision 5.1, 27 July 2006.

10 Thompson, K. Reflections on Trusting Trust. *Communications of the ACM*, 27 (8), 761-763, 2004.

11 Anderson, R. Security in open versus closed systems – the dance of Boltzmann, Coase and Moore. In: *Conference on Open Source Software Economics*, Toulouse, France, 20-21 June 2002, 1-15.

12 Ballard, M. *Google plays cat and mouse with regulators*. The Register, published 2007-05-25. URL: http://www.theregister.co.uk/2007/05/25/google_privacy/

13 Tews, E, Weinmann, R-P, Pyshkin, A. *Breaking 104 bit WEP in less than 60 seconds*. Cryptology ePrint Archive, April 2007. Available at http://eprint.iacr.org/2007/120.pdf

14 *Extended Copy Protection*. Wikipedia, Sampled May 2007. Aavilable at http://en.wikipedia.org/wiki/Extended_Copy_Protection

15 Oleshchuk, V A, Haglund, A, Carlsen, U. A New Method to Protect Against Software Piracy. In: Knapskog, S J and Brekne, T (eds). *NORDSEC'98 – the Third Nordic Workshop on Secure IT Systems*, 81-97, 1998.

16 Haglund, M A, Oleschchuk, V A, Sigbjørnsen, S. *Protection of software against use without permit*. 24 July 2001, 21 pages. United States Patent No. US 6,266,416B1.

17 Køien, G M. RFID and Privacy. *Telektronikk*, 103 (2), 77-83, 2007. (This issue)

18 Frawley, W, Piatetsky-Shapiro, G, Matheus, C. Knowledge Discovery in Databases: An Overview. *AI Magazine*, 13 (3), 57-70, 1992.

19 Granmo, O-C, Oleshchuk, V A. Privacy Preserving Data Mining in Telecommunication Services, *Telektronikk*, 103 (2), 84-89, 2007. (This issue)

20 The Logic of Privacy. *Economist*, 4 Jan 2007.

21 Thorogood, R, Brookson, C. Lawful Interception, *Telektronikk*, 103 (2), 33-36, 2007. (This issue)

22 Svendsen, B. The EU Directive on data retention – An end to justify the means. *Telektronikk*, 103 (2), 31-32, 2007. (This issue)

23 Rognsvåg, K. Privacy and Protection of Personal Data. *Telektronikk*, 103 (2), 27-30, 2007. (This issue)

24 EU Commission. *Promoting Data Protection by Privacy Enhancing Technologies (PETs)*. EU Commission, Reference: IP/07/598, May 2007.

For a presentation of the authors, please turn to page 3.

Secure Multi-party Computations and Privacy Preservation: Results and Open Problems

VLADIMIR A. OLESHCHUK, VLADIMIR ZADOROZHNY

Vladimir A. Oleshchuk is Professor of Computer Science at Agder University College, Norway

Vladimir Zadorozhny is Assistant Professor at University of Pittsburgh, USA

In this paper we give a brief overview of the research in the area of secure multiparty computations (SMC) applied to privacy preservation with special focus on the telecommunication systems. We provide classification of the proposed approaches in the telecom context.[1)]

Introduction

People use telecommunication-based applications daily and the system collects a large amount of information related to their activities. Such telecom networks create opportunities for joint cooperative tasks based on computations with inputs supplied by separate users. Moreover, such computations could be performed even among mutually untrusting parties. Since many user inputs are private information reflecting users' daily activities (for example travel routes, buying habits, and so on), secure multi-party computations (SMC) become a relevant and practical approach for dealing with such applications. Many applications can utilize available private data to improve the quality and security of every day life. For example, private data have been used to develop such important applications as traffic jam monitoring, monitoring of elderly people, anti-terror related monitoring of suspects, etc. Meanwhile, with an increased amount of private data collected in telecommunication networks, the privacy concerns become a critical issue. Some questions that should be addressed include how exactly the private data will be used; can the data be misused to invade people's privacy; and "who is watching the watchers".

Many techniques have been adopted for privacy preservation such as k-anonymity [Sw02], data transformation/randomization (for more details see the paper on *Privacy Preserving Data Mining in Telecommunication Services* in this issue [GO07]), etc. In this paper, we focus on techniques based on secure multiparty computations. Secure multiparty computation is a cryptographic technique that enables computations on data received from different parties in such a way that each party knows only its own input and the result of the computations.

Secure Multi-party Computations and Privacy

Many SMC-based solutions can be used to ensure privacy preservation and protection. Informally, they can be described as a computational process where two or more parties compute a function based on private inputs. Privacy in this context means that none of the parties wants to disclose its own input to any other party.

Formally, a secure multi-party computation problem can be formulated as follows: Assume that there are n parties ($n > 1$) that want to perform some computations jointly. It means that each party is willing to contribute some data to these computations. However, each party wants to contribute its input privately; that is, without disclosing their inputs to the other participating parties or to any third party. Generally, this problem can be seen as a computation of a function $f(x_1, x_2, ..., x_n)$ on private inputs $x_1, x_2, ..., x_n$ in a distributed network with n participants where each participant i knows only its input x_i and no more information except output $f(x_1, x_2, ..., x_n)$ is revealed to any participant in the computation [Goldw97]. Historically, the secure multi-party computation problem was introduced by Yao [Yao82] where a solution to the so-called Yao's Millionaire problem was proposed. (The problem that was formulated is about how two millionaires can learn who is richer without revealing any information about their net worth.)

According to theoretical studies, the general SMC problem is solvable based on the circuit evaluation protocol [Goldr04]. At the same time, it was observed that such solutions are not practical from the efficiency point of view. Therefore, finding efficient problem-specific solutions was recognized as an important direction for future research. For that reason, many specific solutions were proposed in the literature recently. Authors considered several problems from different research areas such as information retrieval, computational geometry, statistical analysis, etc.

Initial research on such problem-specific solutions was performed under the assumption of the ideal security model assuming zero information disclosing. However, despite the fact that the specific solutions for problems mentioned above are more efficient than general solutions, they are still expensive and not

1) *This work is supported in part by the Norwegian Research Council under the BILAT project 174958/D15.*

always very usable in practical applications with constrained resources (for instance energy, computational power, or communication broadband). At the same time, many authors started to notice that in some applications users would be willing to accept a reduced level of security if they could achieve sufficient efficiency, especially in the cases where ideal security solutions are not applicable because of unacceptable performance. It has been shown that in this case, the main goal is to achieve efficiency with a sufficient level of security, but not security itself.

This is the main reason why recent research is focused on a new practical paradigm utilizing a security model with some information disclosure. By relaxing the security imposed restrictions, more practical solutions with better performance may be achieved. In the following section, we briefly describe some of such solutions.

A simple example of efficient SMC that illustrates the idea of privacy preserving computations is the secure sum protocol [Clifton02]. Assume that there are n parties $P_0, P_1, ..., P_{n-1}$ such that each P_i has a private data item d_i, $i = 0, 1, ..., n - 1$. The parties want to compute $\Sigma_{i=0}^{n-1} d_i$ privately, without revealing their private data d_i to each other. The following method was presented in [Clifton02] and solves the problem described above, which is also known as a *secure sum problem*. We assume that $\Sigma_{i=0}^{n-1} d_i$ is in the range $[0, m - 1]$ and P_t is the protocol initiator. At the beginning P_t chooses a uniform random number r within $[0, m - 1]$. Then P_t sends the sum $d_t + r (\mathrm{mod} m)$ to the party $P_{t+1(\mathrm{mod} n)}$. Each remaining party P_i does the following: upon receiving a value x the party P_i sends the sum $d_i + x (\mathrm{mod} m)$ to the party $P_{i+1(\mathrm{mod} n)}$. Finally, when party P_t receives a value from the party $P_{t-1(\mathrm{mod} n)}$, it will be equal to the total sum $r + \Sigma_{i=0}^{n-1} d_i$. Since r is only known to P_t it can find the sum $\Sigma_{i=0}^{n-1} d_i$ and distribute to other parties. Figure 1 depicts how the secure sum algorithm operates for the case of four parties P_0, P_1, P_2 and P_3 when P_1 is the protocol initiator.

Some Building Blocks

Almost all solutions reviewed later in this paper are designed as a combination of some basic building blocks or cryptographic techniques. Below we give a brief overview of some of the most popular techniques together with recommendations for further reading.

Yao's Millionaire Problem

This problem is important in the context of data mining and e-commerce, for instance in such applications as online bidding and auctions. It was considered by Yao in [Yao82] and contains a scheme for secure comparison. The scheme assumes that there are two parties, Alice and Bob, where Alice has a number a and Bob has a number b. Both Alice and Bob want to verify whether or not $a < b$ without revealing information about a and b to each other.

Homomorphic Encryption

Homomorphic encryption is a form of encryption that permits performance of a specific algebraic operation (denoted by $\otimes$) on the plain text by performing a (possibly different) algebraic operation (denoted by $\oplus$) on the ciphertext. The homomorphic cryptosystems are used as a basic building block in many secure multiparty protocols. Several such cryptosystems have been proposed in the literature [Ben94, Pai99, NS98]. More formally, let us consider a public-key cryptosystem with the homomorphic property where encryption and decryption are denoted as $E(\cdot)$ and $D(\cdot)$ respectively. It means that there is an operation on encrypted data, denoted as $\oplus$, that can be used to perform summation on the encrypted data without decrypting. Thus, we can find the encrypted sum of encrypted x and y; that is, $E(x) \oplus E(y) = E(x \otimes y)$. Consequently, since

$$E(yx) = E\left(\underbrace{x \otimes x \otimes \cdots \otimes x}_{y}\right) = \underbrace{E(x) \oplus E(x) \oplus \cdots \oplus E(x)}_{y}$$

we are able to multiply encrypted data if only one of the multipliers is unencrypted. Homomorphic cryptosystems proposed in the literature define operation $\oplus$ as modular multiplication while $\otimes$ is defined as modular summation, XOR or modular multiplication. As a simple example of a homomorphic cryptosystem, we can consider the RSA cryptosystem. It is easy to

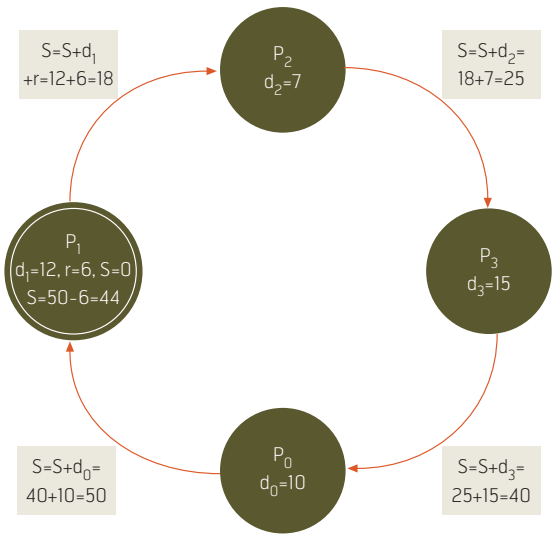


Figure 1 How the secure sum algorithm operates for the case of four parties P0, P1, P2 and P3 when P1 is the protocol initiator

see that $E(x_1) \oplus E(x_2) = (x_1^e \bmod n)(x_2^e \bmod n) = x_1^e x_2^e \bmod n = (x_1 x_2)^e \bmod n = E(x_1 x_2)$, where (e, n) is a public key. In this case, both $\oplus$ and $\otimes$ are modular multiplications. However, in the context of secure multi-party computations, the most used cryptosystems define $\otimes$ as a modular summation [Ben94, Pai99, NS98].

Oblivious Transfer

Oblivious transfer describes communication between two parties, sender and receiver, where the sender transmits part of the data to the receiver. The receiver chooses a part of the received data in a privacy protecting manner: the sender is assured that the receiver gets no more information to which it is entitled. The sender knows nothing about which part of the data has been received. The 1-out-of-n oblivious transfer is a method that allows a party to access one of the n secrets, without getting any information about remaining n - 1 secrets and without disclosing which of n secrets was accessed. The first such protocol was presented by Rabin [Rab81] and since then several variants of different types of oblivious transfer protocols were proposed [EGL85, NP01]. They all serve as important building blocks for many cryptographic applications such as in protocols for signing contracts, certified email or flipping a coin over phone [EGL85]. Theoretically, it was shown by Kilian [Kil88] that by using only oblivious transfer, it is possible to construct any secure protocol. However, oblivious transfer computational requirements are quite demanding in terms of resources and they are often the bottleneck in many applications that invoke them. Therefore, finding efficient solutions is an important research area and many such solutions have been proposed [NP01]. The slightly modified Private Matching example presented in the next subsection can be seen as a 1-out-of-N oblivious transfer protocol.

Private Matching

The objective of the following simple examples is just to give a flavor of how privacy can be achieved. More advanced examples can be found in references.

We assume that Alice has a private set of information $\{a_1, a_2, ..., a_n\}$ and Bob has a private set of information $\{b_1, b_2, ..., b_m\}$. The protocol that follows is a two-party protocol between Alice and Bob that finds the private intersection on their inputs. Alice defines a polynomial $P(x)$ whose roots are her private set A as follows: $P(x) = (a_1 - x)(a_2 - x) \cdots (a_n - x) = \sum_{i=0}^{n} \alpha_i x^i$. Alice sends $E(\alpha_0), E(\alpha_1), ..., E(\alpha_n)$ where $E(\alpha_i)$ is a homomorphic encryption of α_i. Bob evaluates polynomial $P(b_i)$ by finding $E(P(b_k)) = \sum_{i=0}^{n} E(\alpha_i) \cdot b_k^i$ (without knowing the real values of coefficients α_i). After that Bob selects a random number r and calculates $E(r \cdot P(b_k) + b_k)$. If b_k is in A then $P(b_k) = 0$ and $E(r \cdot P(b_k) + b_k) = E(b_k)$. Therefore, Alice can find whether b_k is in intersection $A \cap B$ by decrypting $E(b_k)$ with her private key and checking that $D(E(b_k)) = b_k$ is in A. If b_k is not in A then $P(b_k) \neq 0$ and the result of $D(E(r \cdot P(b_k) + b_k))$ is random. Thus, Alice learns whether b_k is in A without revealing A to Bob and without knowing b_k when b_k is not in A.

Review of SMC-based Privacy Preserving Solutions

The main purpose of this section is to illustrate broad applicability of secure multi-party computations to privacy preservation and to stimulate research on new application areas and problems.

Private Information Retrieval

The problem of private information retrieval (PIR) can be formulated as follows. Assume that a user wants to query a database in a private way, where the database receives no information about the query. Formally, we consider the database as an n-bit string, $x = x_1 x_2 \cdots x_n$ and the query is the bit i. Privacy preserving information retrieval means that the user can retrieve the bit x_i by sending the query i such that the database learns no information about i. The problem was introduced in [CGKS95] and since then studied intensively in literature (see for example [BI01, CG97]). Since the straightforward solution to PIR would be to send the whole database to the user, the main goal of this research was to minimize sublinear communication complexity. However PIR, as it was introduced in [CGKS95], is not concerned with privacy of the database. The extension of the problem introduced in [GIKM98] and called *symmetrically private information retrieval* (SPIR) protects privacy of the database (in addition to user privacy) where database privacy means that the user cannot obtain more information about the database except contained in the result of her query. SPIR can also be seen as very similar to oblivious transfer discussed above.

Selective Private Function Evaluation

This problem was introduced in [CIKRRW01] where several solutions were presented. The problem is formulated as follows: We assume that several servers hold copies of a database $x = x_1 x_2 \cdots x_n$. A client chooses a function f and indices $i_1, i_2, ..., i_k$, and interacts with servers (one or more) in order to compute $f(x_{i_1}, x_{i_2}, ..., x_{i_k})$ privately in the sense that servers know nothing about chosen indices. An example of problem setting that illustrates the usability of selective function evaluation can be described in the following way. Consider a scenario where there is a database containing both public

information about users (for instance their addresses and phone numbers) and private information (for instance age, salary and mobile phone use habits of each user). Public information is freely accessible, but private information is sensitive and should not be accessible. For example, a company wants to perform some statistical analysis on a selected subset of private data without revealing selection criteria while the database owner wants to reveal only data which will be used for analysis.

Privacy Preserving Scientific Computations

Many industries, including the telecommunication industry, have to solve problems related to planning, routing, scheduling or optimization. These problems are often modelled as systems of linear equations or linear least squares problems. However, in the scenario in which two or more untrusted parties want to solve the problems without revealing private data, traditional well-studied approaches are not applicable. Consider for example a scenario in which two telecommunication companies want to optimize joint use of their networks without revealing proprietary information about internal structure, constraints, etc. In many cases, such problems involve solving systems of linear equations where none of the parties has knowledge about every equation. The privacy-preserving case of the problems that have been considered in the literature [DA01] can be formalized in the following way. Assume that Alice has m private equations and Bob has n - m private linear equations represented by $M_A x = c_A$ and $M_B x = c_B$ respectively, where x is an n-dimensional vector. The authors show how Alice and Bob can jointly find a vector x that satisfies all equations without revealing to each other any information about the equations.

Computational Geometry Problems

Several computational geometry problems in the privacy preserving setting have been considered in the literature [AD01]. They include point inclusion in polygons, polygon intersections, finding the closest pair of points, etc. Many of those problems can be easily interpreted in the context of telecom applications [KOle].

The simplest of the computational geometry problems is the so-called Privacy-Preserving Point-Inclusion Problem. The problem can be formulated as follows: Assuming that Alice has a point x and Bob has a polygon P, determine whether x is inside P without revealing to each other any information about the relative position of x with respect to P.

A privacy preserving polygon intersection problem can be formulated as follows. Assume that Alice has a polygon P_A and Bob has a polygon P_B. Both Alice and Bob want to find out whether $P_A \cap P_B$ is empty without revealing any information about the polygons to each other.

Finally, assume that Alice has a set S_A of points in the plane and Bob has a set S_B of points in the plane. The privacy-preserving closest pair problem is about finding (by both Alice and Bob) pairs of closest points among points in $S_A \cup S_B$ without revealing to each other any information about S_A and S_B.

Some Potential Real-life Applications

Secure multi-party computations have a very wide variety of potential applications. As we have already illustrated, many classical computational problems can be reformulated in a privacy preserving manner. In this section, we give examples of some real-life applications that have been considered in the literature.

Privacy Preserving Location

Preserving the privacy of a user is an important challenge for mobile and wireless applications. Informally, it is possible to utilize user location without actually disclosing it either to a service provider or to any third party. In this context we can use the privacy preserving solutions for computational geometry problems [AD01]. In the context of telecommunication systems, the location and identity privacy of the current 2G/3G systems have been analyzed in [KO03a, KO03b, KO05]. Authors argue that using Identity-Based Encryption is ideal for fast set-up in new roaming areas. They proposed solutions for spatial control and location privacy using secure multi-party computations and describe the protocol for privacy preserving based on 3-way authentication and key agreement.

Privacy Preserving Dating System

As a simple example of online collaboration where privacy could be seen as a natural property of the system, we can consider a privacy preserving dating system. In such systems, participants should be able to describe their interests and preferences weighting them by importance. We can look at such a system as a matchmaking process that matches participants by their interests and preferences. Privacy preserving in this context means that the data describing participants will remain private from other participants or the system running the application. A possible approach to implementing such a system is to use private matching and base the private set intersection on it as it is described in [FNP04].

Privacy Preserving Monitoring in Wireless Sensor Network

Consider a wireless sensor network for monitoring vital sign parameters from patients in a metropolitan area. Such a network includes body sensors communicating with a receiver unit carried by a patient, which in turn can use another wireless hop (for example GPRS telecommunication solution) to transfer data to a central base station. Sensor networks transmit monitoring data via a wireless medium and are thus vulnerable in terms of privacy and security. Sensor measurements represent private information about monitored objects, which requires that data transmissions and data flow within and out of the sensor network should be protected. We assume that sensor networks support distributed interaction with the physical environment through measuring and aggregation of data in order to create a dynamic global view. Various streams of measured data can be used to monitor and detect events of interest. Each event is represented as a set of values of monitored parameters.

One approach to protect privacy in such sensor networks would be based on the idea that each sensor node delivers a part of the sensed data, called a share [ZOK07]. Each sensor share is a subset of monitored parameters assigned to that sensor. For example, we can define a function *Share* that maps individual sensors to a power set of monitored parameters. Thus, in order to obtain complete information about the monitored environment (status information), a base station should collect shares from all sensors in the network. The shares should be selected in such a way that individual sensor outputs are not sufficient for reconstructing complete status information. Intelligible reconstruction of the status information is only possible when a certain number N of distinct shares is available, where N is called an intelligibility threshold. For example, assuming that each share is associated with one monitored parameter for each sensor in a sensor network, complete status information includes knowledge of all sensor parameters. The complete status information is associated with the lowest security and privacy requirements, since it reveals all the data delivered by sensors.

Privacy Preserving Electronic Surveillance

In [FA03] the authors consider how to collect data about conditionally dependent individuals whose surveillance is authorized. The authors consider how to monitor activities of only those individuals whose surveillance is authorized without disclosing the identities of monitored individuals to the data collecting entity, while ignoring individuals whose monitoring is not authorized. The authors assume that the set of all identities *U is a subset of* the set *S* of identities for which monitoring is authorized. Let Alice be the monitoring agency that knows *S*, and Bob be a data-collecting entity that can collect data about activities of elements from *U* that he observes. The privacy preservation in this setting means that Alice can learn about activity of identity *p* from *U* if and only if *p* is also in *S*, but collecting entity Bob cannot learn anything about the membership of *p* in *S*. In [FA03], the authors propose protocols to solve this problem.

Privacy Preserving Credit Checking

A privacy preserving credit checking problem was considered in [FAZ05]. It deals with the process of applying for a loan that can be described as follows: Assume that Bob wants to borrow money from a lender, Linda. Before giving a loan to Bob, Linda checks Bob's credit history to find out whether Bob is trustworthy and capable of paying the loan. To do a credit check on Bob, Linda requests a credit report about Bob from a Credit Report Agency. Linda determines if Bob qualifies for the requested loan based on her qualification conditions. As we see from this description, some private information will have to be revealed during this process. For example, Bob's financial information such as borrowing history, spending habits, etc., are commonly described in the credit report. Meanwhile, Linda does not need to know all information from the credit report. What she really needs is to check whether her qualification conditions are satisfied. However, in many cases these conditions are also private and Linda will not reveal them to anybody including a Credit Report Agency. Thus, the privacy-preserving solution of credit checking would involve approving a loan application as it is described above where both the borrower's (Bob) private information from the credit report and the lender's (Linda) qualification criteria remain private. Formally, the problem is defined as follows. We assume that a credit report is presented as a set of attributes $a_1, a_2, ..., a_m$ where a_i is either Boolean or an integer, and the qualification criteria $c_1, c_2, ..., c_n$ where each criterion c_i is a function on a subset of the attributes. The lender's qualification policy that determines whether a borrower qualifies for a specific loan is defined based on what combination of criteria $c_1, c_2, ..., c_n$ are satisfied on $a_1, a_2, ..., a_m$. The proposed solution utilizes secure multi-party computations that solve this problem efficiently. It requires communication and computation resources proportional to the size of the credit report and lender's policy.

Conclusion

In this brief review of the existing research, we considered the main idea behind secure multi-party computations and how they can be used to develop distributed applications that preserve the privacy of participating parties. By using simple examples, we

demonstrated how SMC techniques have been used to develop novel applications with privacy preservation as an essential property. However the most common drawback of SMC protocols, which substantially impacts their applicability, is their inefficiency. They require both considerable computational and communicational resources. Not many, if any, of such applications are implemented. However, as both the availability of such resources and privacy concerns are continuously growing, one would expect that many such applications may be implemented in the near future. An important problem is whether *practical* solutions exist that are based on an ideal security model. Therefore, future research should seek other security models that can provide low-cost practical solutions with acceptable security level for given type of applications [DZ02]. Finding such practical solutions that balance security and efficiency is an important research area.

References

[AD01] Atallah, M J, Du, W. 2001. Secure Multi-party Computational Geometry. In: Dehne, F K, Sack, J, Tamassia, R (eds). *Proceedings of the 7th international Workshop on Algorithms and Data Structures* (8-10 August 2001). Lecture Notes In Computer Science, 2125. London, Springer-Verlag, 165-179.

[BI01] Beimel, A, Ishai, Y. 2001. Information-Theoretic Private Information Retrieval : A Unified Construction. In: Orejas, F, Spirakis, P G, Leeuwen, J v (eds). *Proceedings of the 28th International Colloquium on Automata, Languages and Programming*, 8-12 July 2001. Lecture Notes in Computer Science, 2076. London, Springer-Verlag, 912-926.

[Ben84] Benaloh, J. Dense Probabilistic Encryption In: *Proceedings of the Workshop on Selected Areas of Cryptography*, Kingston, ON, May 1994, 120-128.

[CIKRRW01] Canetti, R et al. 2001. Selective private function evaluation with applications to private statistics. In: *Proceedings of the Twentieth Annual ACM Symposium on Principles of Distributed Computing, PODC '01*, Newport, Rhode Island. New York, NY, ACM Press, 293-304.

[CG97] Chor, B, Gilboa, N. 1997. Computationally private information retrieval (extended abstract). In: *Proceedings of the Twenty-Ninth Annual ACM Symposium on Theory of Computing*, El Paso, TX, 4-6 May 1997, STOC '97. New York, NY, ACM Press, 304-313.

[CGKS95] Chor, B, Golrreich, O, Kushilevitz, E, Sudan, M. Private Information Retrieval. In: *Proceedings of IEEE Symposium on Foundation of Computer Science*, USA, 23-25 October 1995.

[Clifton02] Clifton, C et al. Tools for privacy preserving distributed data mining. *J. SIGKDD Explor. Newsl.*, 4 (2), 28-34, 2002, ACM Press.

[DuAt01] Du, W, Atallah, M J. Secure multi problem computations problems and their applications: a review and open problems. *NSPW'01*, 12-13 September 2002, 13-21.

[DA01] Du, W, Atallah, M J. 2001. Privacy-Preserving Cooperative Scientific Computations. In: *Proceedings of the 14th IEEE Workshop on Computer Security Foundations, CSFW*, 11-13 June 2001. Washington, DC, IEEE Computer Society, 273.

[DZ02] Du, W, Zhan, Z. 2002. A practical approach to solve Secure Multi-party Computation problems. In: *Proceedings of the 2002 Workshop on New Security Paradigms, NSPW '02*, Virginia Beach, VA, 23-26 September 2002. New York, NY, ACM Press, 127-135.

[EGL85] Even, S, Goldreich, O, Lempel, A. 1985. A randomized protocol for signing contracts. *Commun. ACM,* 28 (6), 637-647, 1985.

[GO07] Granmo, O-C, Oleshchuk, V. Privacy Preserving Data Mining in Telecommunication Services. *Telektronikk*, 103 (2), 84-89, 2007. (This issue)

[FNW96] Fagin, R, Naor, M, Winkler, P. 1996. Comparing information without leaking it. *Commun. ACM*, 39 (5), 77-85, 1996.

[FNP04] Freedman, M J, Nissim, K, Pinkas, B. 2004. Efficient Private Matching and Set Intersection. In: *Advances in Cryptology – EUROCRYPT 2004*. Lecture Notes of Computer Science, 3027. Springer-Verlag, 1-19.

[FA03] Frikken, K B, Atallah, M J. 2003. Privacy preserving electronic surveillance. In: *Proceedings of the 2003 ACM Workshop on Privacy in the Electronic Society, WPES '03*, Washington, DC. New York, NY, ACM Press, 45-52.

[FAZ05] Frikken, K, Atallah, M, Zhang, C. 2005. Privacy-preserving credit checking. In: *Proceedings of the 6th ACM Conference on Electronic Commerce, EC '05*, Vancouver, BC, 5-8 June 2005. New York, NY, ACM Press, 147-154.

[GIKM98] Gertner, Y, Ishai, Y, Kushilevitz, E, Malkin, T. 1998. Protecting data privacy in private

information retrieval schemes. In: *Proceedings of the Thirtieth Annual ACM Symposium on theory of Computing, STOC '98*, Dallas, TX, 24-26 May 1998. New York, NY, ACM Press, 151-160.

[Goldw97] Goldwasser, S. 1997. Multi party computations: past and present. In: *Proceedings of the Sixteenth Annual ACM Symposium on Principles of Distributed Computing, PODC '97*, Santa Barbara, CA, 21-24 August 1997. New York, NY, ACM Press, 1-6.

[Goldr04] Goldreich, O. *The Foundations of Cryptography*, volume 2. Cambridge University Press, 2004.

[Kil88] Kilian, J. 1988. Founding cryptography on oblivious transfer. In: *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing, STOC '88*, Chicago, Illinois, 2-4 May 1988. New York, NY, ACM Press, 20-31.

[KO03a] Køien, G M, Oleshchuk, V A. 2003. Privacy-Preserving Spatially Aware Authentication Protocols : Analysis and Solutions. In: *Proceedings of NORDSEC 2003*, Gjøvik, Norway, 161-173.

[KO03b] Køien, G M, Oleshchuk, V A. 2003. Spatio-Temporal Exposure Control; An investigation of spatial home control and location privacy issues. In: *Proceedings of the 14th Annual IEEE Symposium on Personal Indoor Mobile Radio Communications, PIMRC*, 2760-2764, Beijing, China, Sep. 2003

[KO05] Køien, G M, Oleshchuk V A. 2006. Location Privacy for Cellular Systems; Analysis and Solution. *Proceedings of the 5th Workshop on Privacy Enhancing Technologies, PET'05*, Dubrovnik (Cavtat), Croatia, 30 May - 1 June 2005. Lecture Notes in Computer Science, 3856, Springer, 40-58.

[NS98] Naccache, D, Stern, J. 1998. A new public key cryptosystem based on higher residues. In: *Proceedings of the 5th ACM Conference on Computer and Communications Security, CCS '98*, San Francisco, CA, 2-5 November 1998. New York, NY, ACM Press, 59-66.

[NP01] Naor, M, Pinkas, B. 2001. Efficient oblivious transfer protocols. In: *Proceedings of the Twelfth Annual ACM-SIAM Symposium on Discrete Algorithms*, Washington, DC, 7-9 January 2001. Symposium on Discrete Algorithms. Society for Industrial and Applied Mathematics, Philadelphia, PA, 448-457.

[Pai99] Paillier, P. 1999. Public-Key Cryptosystems Based on Composite Degree Residuosity Classes. In: *EUROCRYPT'99*, Lecture Notes in Computer Science 1592, 223-238.

[Rab82] Rabin, M O. *How to exchange secrets by oblivious transfer*. Aiken Computation Laboratory, Harvard University, 1981. (Technical Report TR-81)

[Sw02] Sweeney, L. 2002. Achieving k-anonymity privacy protection using generalization and suppression. *Int. J. Uncertain. Fuzziness Knowl.-Based Syst.*, 10 (5), 571-588, 2002.

[Yao82] Yao, A C. Protocols for secure computations. In: *Proceedings of the twenty-third annual IEEE Symposium on Foundations of Computer Science*, IEEE Computer Society, 1982, 160-164.

[ZOK07] Zadorozhny, V, Oleshchuk, V, Krishnamurthy, P. A Framework for Efficient Security and Privacy Solutions in Data Intensive Wireless Sensor Networks. *Telektronikk*, 103 (2), 61-76, 2007. (This issue)

For a presentation of Vladimir A. Oleshchuk, please turn to page 3.

Vladimir Zadorozhny is an Assistant Professor in the Department of Information Science and Telecommunications, University of Pittsburgh. He received his PhD in 1993 from the Institute for Problems of Informatics, Russian Academy of Sciences in Moscow. Before coming to USA he was a Principal Research Fellow in the Institute of System Programming, Russian Academy of Sciences. From May 1998 he worked as a Research Associate in the University of Maryland Institute for Advanced Computer Studies at College Park and joined University of Pittsburgh in September 2001. His research interests include networked information systems, wireless and sensor data management, query optimization in resource-constrained distributed environments, and scalable architectures for wide-area environments with heterogeneous information servers. Dr. Zadorozhny has served on program committees of multiple Database and Distributed Computing Conferences. He also co-chaired the technical program of MDDS 2005 and DISN 2006. He has delivered several tutorials in conferences and at universities abroad, of which the most recent was on Network Aware Wireless Sensor Data Management 7th International Conference on Mobile Data Manaagement, Nara, Japan, 2006.
email: vladimir@sis.pitt.edu

Privacy and Protection of Personal Data

KJETIL ROGNSVÅG

Kjetil Rognsvåg is Manager of Telenor Privacy Office

The telecom industry has more information about their customers and employees than most other businesses. With a growing concern in the public rearding the use and misuse of personal data, and with political focus on extended data retention in our business, we need to be more aware than ever before of how we handle our data.

The European Convention for the Protection of Human Rights and Fundamental freedoms states that "Everyone has the right to respect for his private and family life, his home and his correspondence" [1].

If you ask people around you, I guess most of them also would agree to the 'definition' in the Personal Data Act [2] which answers: "To protect natural persons from violation of their right to privacy through the processing of personal data" and to "... ensure that personal data are processed in accordance with fundamental respect for the right to privacy, including the need to protect personal integrity and private life and ensure that personal data are of adequate quality".

The borderlines for what constitutes "adequate quality" and "fundamental respect" of private life will, however, vary quite a lot depending on customs, nationality, political preferences and economic incentives.

Privacy has for decades been a subject in Telenor. The 'ladies at the telephone exchange' could be fired if they listened in on a conversation. Today, all employees have signed the "Codes of conduct", where paramount requirements for protection of personal data are stated. Furthermore, all employees have signed the declaration of confidentiality where we are reminded of our duty not to give out any information regarding our customers we might have obtained at work. It is considered a severe offence if an employee violates that rule.

Telenor has furthermore secured the access to data and to systems containing personal data, and as part of the Norwegian total defence we even made the security around our systems much higher than one would normally expect from a telecom company.

Over the years Telenor has handled various products to even further secure customers' extra need for privacy, with unlisted numbers and reservation services, according to both legislation and good customer care.

The new Personal Data Act from April 2000 came into force 1 January 2001. Rules of transition allowed companies to continue as before until 1 January 2003. At the beginning of 2002, Telenor escalated the work on mapping how we handle personal data. A larger project on Telenor group level followed several smaller projects in Telenor companies where we tried to sort out which requirements should be adapted and how, and if there was any derogation between the ways the companies handled data and the interpreted legislative acts.

As a continuation of these projects, Telenor Privacy Office was established as a network organisation in October 2004, with a mandate to establish routines ensuring that Telenor complies with the relevant demands on this subject. The Privacy Office covers all units and employees who handle personal data in Norway and in the Nordic countries.

I will now try to give my view on some aspects related to privacy and our business.

Privacy in Telecom Services and Our IT Solutions

One of my personal favourite stories when it comes to showing the importance of safeguarding the customers' privacy in secure telecom services is the poor guy who used the camera on his mobile to photograph his 'private parts'. The picture was only intended for his girlfriend, but due to an error in the telecom systems, the MMS with the picture was sent to several hundred mobile subscribers. Just minutes after he sent the message, furious people started to answer him back.

As Privacy Ombudsman in Telenor, I have also received anxious phone calls from customers who wonder who has access to the content of their MMS, SMS and e-mails.

Two years ago, Telenor was in the media for violating our customer's privacy: "Anyone may check any customers' bill", was the headline. The criticisms were mainly that an automated service made it possible to check the outstanding amount on the bills,

 ISSN 0085-7130

without any precautionary measures whatsoever. Telenor immediately changed the service so the caller also had to input his birth date – this was not sensitive data at all, and the security level therefore could be set quite low.

At about the same time, we had another challenge with our routines for unlisted numbers: According to the Electronic Communications Regulation [3], we have to provide unlisted phone numbers for customers who require such. A person with an unlisted number updated her home address – we need that for billing purposes. Unfortunately, the address connected to an older subscription of hers with another number was updated as well. This subscription had not been used for quite some time, and the owner had therefore no thought of marking this as an unlisted number, too. It was simply not used. The result was that the address of this customer was displayed on the older number at the information services and in the public directory.

This incident led to changes in our routines and also shows how important it is to think of data security and privacy when we develop routines and systems to support the routines.

Securing privacy is not only about IT solutions! No matter how secure the IT system itself is, the routines for the people handling the systems may very well be the weak link. Security does not give you privacy, even though you depend on security to have privacy.

Our Vast Amount of Personal Information

Information used for marketing and sales activities are mainly fresh data, that is not more than 3-6 months old, and for some usage up to one year.

In Norway, we are furthermore allowed to keep data about a former customer for up to two years after he has left us, but we are not allowed to use the data during that time for marketing purposes. Still, when we performed a review of all our IT systems and databases two years ago, we discovered that very few of the systems had good routines for deletion of old data. Sad to say we found that a common deletion 'routine' was "... the hard disk is almost full; we have to get rid of something here ...". No wonder that over the last years we have spent quite some resources to automate new routines.

Traffic data on 'invoice level' is kept for 10-11 years due to legislation. For the same reason detailed level traffic data so far is kept only for 3-5 months for fixed, mobile and IP telephony. For Internet traffic we have kept information from some days up to some weeks.

Now the new EU directive on data retention [4] will change this dramatically – not only will the retention period increase, but also the amount of different kinds of data to be stored will be magnified. In order to fight organized crime and terrorism, service providers of telecom- and other electronic communication services have to store data for a period of between six and 24 months, according to the national implementation of the directive.

The vast quantity of detailed personal data we possess will of course be regarded as a possibility for some, and a threat for others. And the threat is certainly not only for the criminals! The various levels of quality of the data from the different service providers represent a risk for all citizens in Europe. I know that small errors may cause wrong use of data, like sending out wrong data or replicating wrong data from one system to another. I know that Telenor put up quite a lot of resources to achieve sufficient data quality in order to use the data for various purposes.

And I wonder: Will an office outside of the service providers have enough skill about the information to know each and every weakness in data quality from the different service providers? And when we – as a service provider with very good knowledge of our own customer data, data quality and routines – still makes mistakes, what are the chances that an office far away from the business, possibly with data from between 200 – 300 service providers for Norway alone will make less mistakes? And which consequences would an error from such an office have? You might not be convicted of a crime – but what about the consequences of being suspected of a crime?

For Telenor and other e-com service providers I see no advantages at all in expanding the retention period. When it comes to advantages for the European society as a whole, the answer depends on who you ask.

If I were to fight terrorism and organized crime and were to spend billions of Euros on the attempt, I for sure wouldn't put all the money on data retention in the European e-com industries. For well-organized crime there still seem to be quite a few ways to communicate without leaving traces, and the billions of Euro would be better spent on other activities. But obviously, it is easier to introduce longer retention periods which the service providers and end-users of the services will have to pay, rather than increasing government spending on for example police, investigations and prosecution.

Privacy Versus Marketing

"Hi, would you like to save thousands on your mobile phone bill?" Even I, as a Telenor employee, have received calls trying to convince me that I would be far better off if I moved my subscription to them ... In that case, they obviously had very poor data quality in their call lists ...

But marketing and sales people have and should of course have a strong focus on their goals to succeed. The goals may be customer acquisition, subscriptions sold, increased use of a service etc. Without the proper incentives, guidelines and requirements, it will be tempting to use 'any' means to reach the goals.

We, like many other companies, have experienced employees and consultants who, in order to reach their goals, sometimes went a bit over the standards we try to live by. Telenor continuously works to ensure that we follow our guidelines and follow up when someone cross our ethical, and sometimes even legal borders. This is extremely important when you think of the vast quantity of personal data we possess.

When the ministry of transportation distributed a questionnaire during the spring 2005, the sad result showed that the telecom companies were the least trusted group when it came to handling personal data. The groups which followed were the toll-road companies, and non-profit organisations. My thought – and hope – is that the result was partly due to bad behaviour and aggressive marketing from some of the new companies which were established around 2004 – companies that are luckily no longer in the market.

The data we have could furthermore very well be used to build really detailed profiles of a person, so we then could send out very adapted messages, probably with a very good return on investment. To a certain extent Telenor and other companies do adapt messages to certain groups, namely based on age, type of product and service, and to a certain extent whether a customer uses a service or not.

But most of us do not want our vendor of telecom services to be too nosy about where we are located, whom we are talking to or which Internet sites we are visiting. This is our private sphere, and even though I know that my vendor has these data, I would be more than disappointed if I discovered that they really dug into my private life just to sell me more bandwidth!

It is however considered good customer care to present relevant information, instead of bombarding people with junk-mail. Actually this was the very background for the establishment about ten years ago of the Data Quality Manager role in marketing in Telenor Mobil. At that time the then CEO had been declared the king of junk-mail in Norwegian newspapers, a title he really disapproved of.

The other day a marketing consultant told me that he saw that I had ADSL, and "wouldn't I like to have IP-telephony?" How could he know, and what else does he know about me? In this case he didn't know anything at all; he guessed, but our average customers start wondering where the data flow. They are afraid that information they give away will result in more spam in their mailboxes, or that their privacy is disturbed by more calls from people selling services. Many people do realize that personal data is big business.

Consent and Reservations

As indicated above, the hard competition is continuously driving the companies to set out even more creative ways to reach their sales goals. There has been focus on aggressive telemarketing in the press several times over the last years, where people have questioned where the marketers have got their information from, and why the various reservations and need for consent are not respected.

In Norway, we have for some years now had an official reservation register in Brønnøysund [5] for direct marketing by mail or telephone. In spite of this register people still receive unsolicited marketing. One of the explanations is simply that they do not know that as long as you have a customer relationship with a company, the company is allowed to do direct marketing towards you without checking the register in Brønnøysund. However, the company itself has to have a reservation register where their own customers can make their reservations.

Another explanation is different spelling of names in customer registers and in the register in Brønnøysund, since comparison of names and birth dates so far is the only way to filter out persons who have reservations – we are not allowed to use common customer identification like the social security number.

At Telenor, own reservation register was also a challenge due to our history: We were initially divided into three separate companies for the fixed line, mobile and Internet services; each of them with their own reservation registers. Early last year we introduced one single reservation register for the consumer market, giving us and the customers far better control of the reservations.

I am happy to say that we have received positive response for our various efforts to handle personal information in a better way, also from the Norwegian Data Inspectorate (Datatilsynet) [6].

Privacy for Employees

As an employer of several thousand persons, Telenor also have more data on most employees than many other companies. And because of the focus on data security and privacy in Telenor, many employees do have a special concern for their own privacy as well.

Several cases in the news during the last years have given employees an understanding of the possibilities for misuse of data. It more or less started with the scandal in the former Finnish telecom operator Sonera, where people in the top management were arrested and later got a suspended sentence to prison for having tapped conversations to find out whom were to blame for leakage of information. The police do have that possibility under strict regulation [7], but tapping a conversation is not for everybody!

In Norway we have had various cases regarding the employers looking into employees' e-mails, also here with a very strong presentation of the stories in the media.

Even though Telenor still have, and probably always will have improvements to do when it comes to routines and guidelines for handling employee data, I'm glad to see that the press in this matter have used us as an example of a company with sufficient and good documentation and handling.

Finally, one of the most important things to remember, both as an employer, and as a service provider, is the rule of always informing the registered person about what kind of data you register, and what you are going to do with the data. You are obliged to have a valid cause for handling personal information, and if you can't find such a cause in the privacy act, you simply cannot handle the data in that way!

References

1 Council of Europe. *Convention for the protection of Human Rights and Fundamental Freedoms.* Article 8 – Right to respect for private and family life. http://conventions.coe.int/Treaty/en/Treaties/Html/005.htm

2 Norwegian Act of 14 April 2000 No. 31 relating to the processing of personal data (*Personal Data Act*), section 1, implementing the Directive 95/46/EC of the Eiropean Parliament and of the Council in Norway. http://www.datatilsynet.no/upload/Dokumenter/regelverk/lov_forskrift/lov-20000414-031-eng.pdf

3 *Norwegian Electronic Communication Regulations* § 6-2. More or less the same formulations can be found in Directive 2002/58/EC of the European Parliament and of the Council, Art. 12, paragraph 2. http://www.lovdata.no/for/sf/sd/sd-20040216-0401.html

4 Directive 2006/24/EC of the European Parliament and of the council of 15 March 2006.

5 Brønnøysundregistrene. Available at http://www.brreg.no/

6 Datatilsynet. Available at http://www.datatilsynet.no/

7 Thorogood, R, Brookson, C. Lawful Interception. *Telektronikk*, 103 (2), 33-36, 2007. (This issue)

Kjetil Rognsvåg is manager of the Telenor Privacy Office, and since October 2006 also Privacy Ombudsman for Telenor ASA. He has a B.Sc. in electronics, and has several years of experience from IT and marketing departments, quality assurance and project management, mainly at Ericsson in Sweden and Norway, and from various companies within the Telenor group. Six years ago he started to dig into the subject 'privacy', and the last three years this has been his full time occupation.

email: kjetil.rognsvag@telenor.com

The EU Directive on Data Retention – An End to Justify the Means

BERIT SVENDSEN

Berit Svendsen is VP Telenor Nordic Fixed

Communications traffic data are essential for law enforcement agencies when investigating serious crime and terrorism. Following the terrorist bombings in Madrid and London the EU felt a strong need for harmonised rules on communications data retention throughout the Union. A time-consuming process in order to establish such rules resulted in a Directive on Data Retention which was finally adopted by EU in February 2006. The Directive contains detailed rules forcing telecom operators to retain call and Internet records for use in anti-terror investigations. Member States must implement the Directive in national law within August 2007. Main points of the rules and their impact on telecom oeprations are discussed in the article.

After the Madrid bombings in 2004, a meeting of EU Member States led to the publication of the European Council's Declaration on Combating Terrorism[1]. This document specifies Member States' intention to draft proposals for rules relating to the retention of communications traffic data by service providers. The rules and regulations ensuring that national authorities have legal access to such traffic data vary greatly between Member States, and the EU therefore considered it urgent to establish harmonised regulations throughout the Union, as a means to ensure that efforts to combat the increasing terrorism in Europe are effective. Following discussions of the proposals and the legislative process, a new Directive on Data Retention was finally adopted by the EU in February 2006[2].

The EU Directive on Privacy[3] contains harmonised rules relating to the protection of personal data whenever traffic data are processed in relation to the use of electronic communications services. Such traffic data should be deleted or made anonymous when they are no longer needed for the effectuation of communication, or for invoicing purposes. But access to such traffic data is important to ensure proper identification of subscribers and users of services whenever this is needed for purposes relating to law enforcement and security, such as prevention and prosecution of serious crime and terrorism. This is recognised in the Directive on Privacy and Electronic Communications, which opens for a softening of the restrictions relating to data protection whenever this is required to ensure appropriate safeguarding of national and public security and the prevention of crime. To ensure that harmonisation of provisions against terrorism is actually achieved, the Directive on Privacy has been amended. Requirements relating to the protection of personal data may thus be overruled whenever this is deemed necessary to combat crime and terrorism.

The EU Member States are obliged to incorporate the new Data Retention Directive in their national legislation. In accordance with the EEA agreement, this obligation is also applicable to Norway.

The directive applies to providers of publicly available electronic communications services and public communications networks. This covers a large number of companies, such as fixed-line and mobile telecommunications operators, satellite operators, cable operators, Internet service providers and companies that provide electronic communications services such as web mail, instant messaging or voice over IP.

The directive harmonises the obligations to retain traffic, location and identification data that operators generate or process when supplying communications services. The data to be retained include both successful calls and unsuccessful call attempts, to allow the tracing and identification of the source and destination of a communication, the date, time and duration of communication, the type of communication and the equipment used for the communication and its location. For fixed-line telephony, retained data will include the telephone numbers of callers and those receiving the calls, but also numbers involved in rerouting, names and addresses of subscribers or registered users, and the type of telephone service used. For Internet services, data revealing users' identities (ID) and IP addresses must be retained. Data

1) *Declaration on combating terrorism, European Council, 25 March 2004.*

2) *Directive on the retention of data processed in connection with the provision of public electronic communication services and amending Directive 2002/58/EC.*

3) *Directive 2002/58/EC on Privacy and Electronic Communications.*

 ISSN 0085-7130

related to mobile services must also include the international mobile subscriber identity of callers and those receiving the calls (IMSI) and the international mobile equipment identity of both parties (IMEI).

All categories of data covered by the directive must be retained for a minimum of six months and a maximum of two years. Member States are free to set this period within these limits. In exceptional circumstances, Member States can allow retention periods of more than two years for a limited time, subject to the Commission's approval.

Member States must incorporate the directive in national law within 18 months of its adoption; that means until August 2007. However, another 18 months can be admitted for the implementation of the provisions on Internet access, Internet telephony and Internet e-mail. Before this becomes effective in Norway, both the obligations and the implementation must formally be approved by the EEA Committee. However, a public commission has been appointed by the Norwegian government to consider the obligations and likely national options.

Telecoms operators in Europe, including Telenor, are sceptical to the retention of such huge data and the individual registering of any customer's use of electronic communication, fixed-line telephone services, mobile services and the Internet. The new requirements are considerably more comprehensive than the measures which are practised in Norway and most European countries, and represent huge challenges relating to the protection of personal data and security. There is a need for new data storage systems and for systems which can give legal authorities access to the required data. Furthermore, the EU has left it to the Member States to decide how to cover the considerable costs that the retention requirements involve for the operators. History shows that governments, at least the Norwegian authorities, prefer to impose costs relating to the implementation of requirements from governmental authorities on the operators. This will necessarily result in increased prices on telecoms services. Much could be gained from European standardisation of technology and procedures in this area, but so far the EU has not taken any initiatives to encourage such standardisation.

If the EU's purposes are to be fully achieved, the obligation for data retention must also embrace companies and institutions which are not normally considered to be providers of public electronic communications. Such companies include hotels, Internet cafes, universities and companies which allow a number of people to use their communications services. Hotels offering such services to guests would be obliged to register and store data relating to each guest room, together with identification data on the guests. Similar obligations would be put to Internet cafes, academic institutions and public institutions which offer the public access to their communications facilities. So much storage of personal data, with so many players, also means an increased risk of compromising sensitive personal data.

Telecommunications is one of the most dynamic industries in the world today. New global players like Google, Yahoo, Skype and Microsoft offer communications services based on the Internet. Skype offers opportunities to make telephone calls via the Internet from PCs, and 130 million customers all over the world are now using this option. The new Internet message service, Instant Messaging, has grown tremendously during the last 2-3 years, and the service is currently used by several hundred million people. It is not likely that the EU requirements on data retention will encompass any of these global providers.

If global providers and national players, which for some reason are not considered to be first-hand providers of public electronic communications services, are not required to retain traffic data, it is likely that dangerous criminals will be among the first to detect the gaps in the system and make use of them. If that were to happen, the EU measures will have a minimal effect on combating terrorism.

Berit Svendsen holds an MSc in Electronics from the Norwegian University of Science and Technology (NTNU) (1988), and a Master of Technology Management from NTNU and Massachusetts Institute of Technology, USA (1995). She joined Telenor in 1988 as Research Scientist, and was Director of Telenor's Fixed-Mobile Convergence project 1999-2000. Berit Svendsen became Head of Department and subsequently Managing Director of Data Services in Telenor Network AS. From 2000 to 2005 she was Executive Vice President and CTO of Telenor and also working Chair of Telenor R&D. In January 2005 she took up the position as Vice President of Telenor Nordic Fixed with overall responsibility for the fixed network in Norway. Her special areas of interest are convergence of telecommunications and internet, and industrial development and innovation. From 2002 to 2007 Berit Svendsen was a member of the European Commission / IST Advisory Group. She has a seat on the Board of Ignis AS. Berit Svendsen has held a series of international and national presentations.
email: berit.svendsen@telenor.com

Lawful Interception

RUPERT THOROGOOD, CHARLES BROOKSON

Rupert Thorogood is a Consultant with the Home Office under the UK government working on lawful interception

Charles Brookson is Assistant Director in the Department of Trade and Industry, UK

In order to be effective Lawful Interception facilities must be built within new technologies from the start of the design and standardisation. The European Telecommunications Standards Institute (ETSI) has for over ten years successfully defined the various interfaces, and provided the facility in systems such as mobile and Next Generation Networks. This paper gives an overview of the various standardisation initiatives, and the technologies to which it has been successfully deployed.

What is Lawful Interception?

Lawful interception (LI) is the legally authorised process by which a network operator or service provider gives law enforcement officials access to the communications (telephone calls, e-mail messages etc) of private individuals or organisations. Lawful interception is becoming crucial to preserve national security, to combat terrorism and to investigate serious criminal activities.

The work in Lawful Interception has its foundation in the European Council Resolution of January 1995 [29] which outlined the International Requirements for the Lawful Interception of Telecommunications now known widely as the IUR. This was the result of several years of work by the European governments in cooperation with Australia, New Zealand, Canada and the USA.

The standardisation of lawful interception is vital to provide an economically and technically feasible solution that complies with national and international conventions and legislation. ETSI has played a leading role in the standardisation of lawful interception since 1991; today work is concentrated in Technical Committee Lawful Interception (TC LI), which enjoys the active participation of the major telecom manufacturers, network operators, Law Enforcement Agencies and regulatory authorities of Europe and from around the world.

ETSI's LI work covers the whole spectrum of interception aspects, from a logical overview of the entire architecture and the generic intercepted data flow, to the service-specific details for e-mail and Internet.

Achievements

A major achievement of ETSI's work in this area has been publication of the specifications for the handover procedure: TS 101 671 [6] and ES 201 671 [1]. These specifications illustrate the flow that the intercepted data should follow in telecommunication networks or services. In this context, they specify the network or service protocols necessary to provide lawful interception, as well as the physical or logical point at which the interception has to take place (the handover interface) both for packet data and switched-circuit communications. Other relevant specifications cover LI network functions [2] and various technical specifications relating LI services in ISDN networks and IP-based services etc [7-12].

ETSI has also produced other important specifications on lawful interception in other Technical Committees. For this reason, TC LI is working in close collaboration with TC TISPAN, the Committee in charge of creating the specifications for Next Generation Network (NGN) in ETSI as well as with other relevant committees (TC TETRA, 3GPP and TC Access and Terminals (TC AT) [13-28].

The LI handover specifications are already widely used. They were first adopted in 2003 by the Netherlands regulation authority (Directorate General for Telecommunication and Post of the Ministry of Economic Affairs). Meanwhile a number of other countries are in the process of implementation or have expressed an interest in adopting the specifications.

The specifications are subject to constant review and updating within ETSI to accommodate emerging needs, and are being used as the basis for specifying the procedures for lawful interception. The increasing trend in the use of packet-switched technologies has necessitated a standard for the delivery of IP-based interception: TS 102 232-1 [3] specifies the approach, the protocols and headers needed to perform lawful interception on an IP-based platform.

In addition, lawful interception has to be possible on specific services (Service-Specific Details, SSD) that make use of the IP framework: TS 102 232-2 [4] covers the service-specific details for e-mail services, describing the handover to the law enforcement authorities, whilst TS 102 233 [5] covers the service-specific details for Internet access.

 ISSN 0085-7130

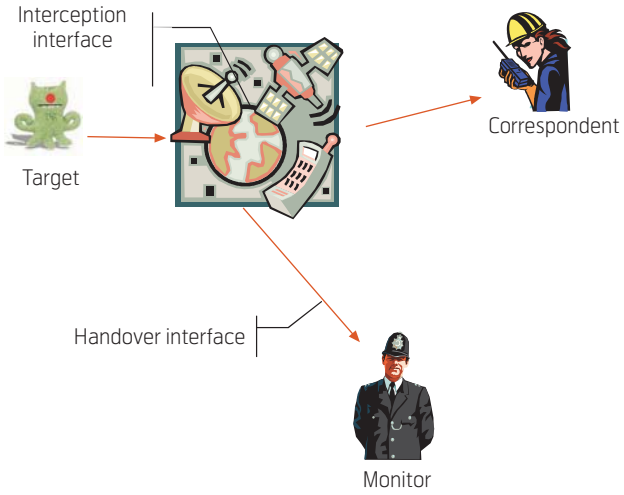


Figure 1 Generalised interfaces for LI specifications

A simple architecture for LI is shown in Figure 1; ETSI specifies the Interception and Handover interface.

ETSI has also standardised the general requirements of network operators, service providers and access providers who are obliged to make available results of interception to the law enforcement agencies. Complementing this, a Technical Specification (TS) relating to handover interfaces for the interception provides guidance for law enforcement agencies on the co-operation required by network operators/service providers with the lawful interception of telecommunications (Figure 2).

Recent publications include a specification on service-specific details for layer 2 lawful interception. This specification applies to access providers having access to information on layer 2 session information. This TS is particularly important because, in many situations, information on higher layers is either not accessible or not stored.

Figure 3 summarises the deliverables produced and their placement in the overall architecture for lawful interception in relation to the ISO-OSI protocol stack.

Ongoing Activities

A specification on the lawful interception of public Internet access by means of wireless LAN technology is being produced. This is a critical issue for lawful interception because the user cannot always be identified.

IP Multimedia subsystem (IMS), the system created in 3GPP to enable the provision of multimedia services, and TISPAN specifications are being developed in tandem to allow the convergence of fixed and mobile networks over this common IP-based platform. The handover interface for lawful interception is being developed in TC LI to align with the latest TISPAN and 3GPP specifications for NGN.

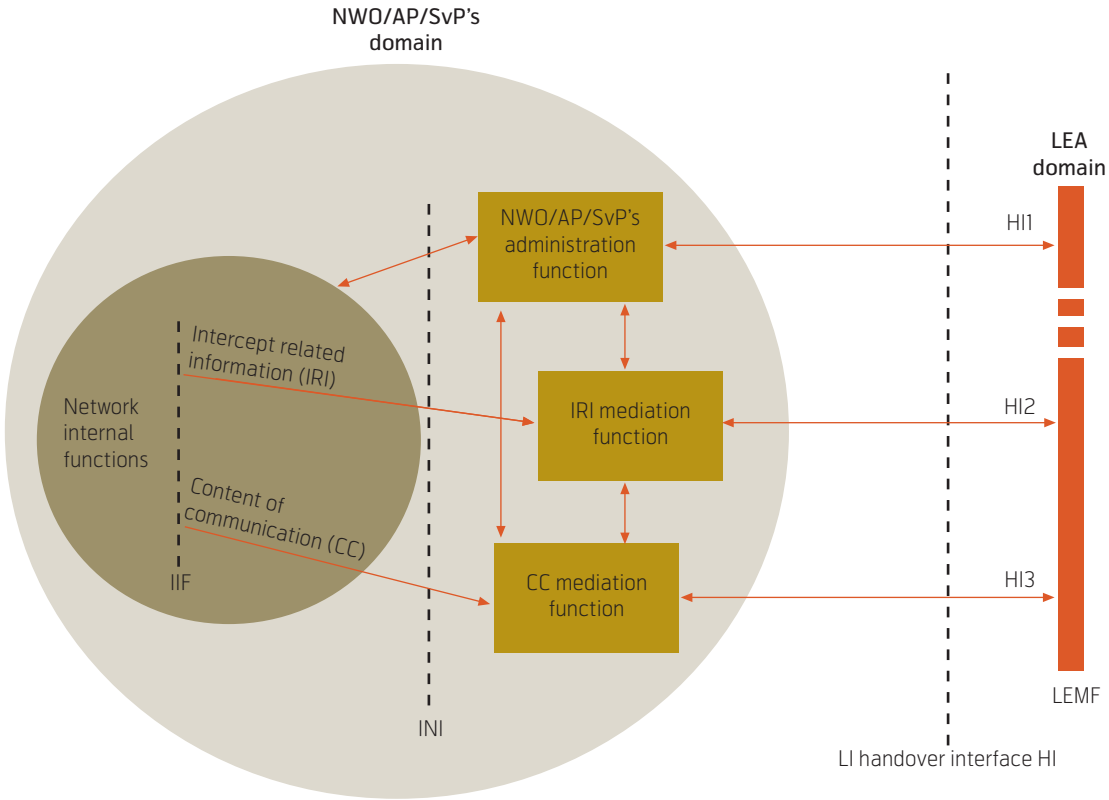


Figure 2 Showing the interfaces between Operators and Service Providers and the Handover (HI) interface

TC LI is also addressing Data Retention. European governments are becoming increasingly interested in preserving communications related information. The European Parliament's civil liberties committee recently voted in favour of new rules, whereby details on telephone calls and Internet use would be kept for six to 12 months. TC LI is producing a specification (DTS/LI-00033) which will provide a handover interface for the request and delivery of retained data between government Agencies and providers of communication services or their agents, based on common global capability needs.

Conclusions

The ETSI specifications provide a building block for present and future architectures of communications standards, which are supported by many manufacturers and providers of services. The main challenge is to continue to evolve the specifications to support new services and technologies!

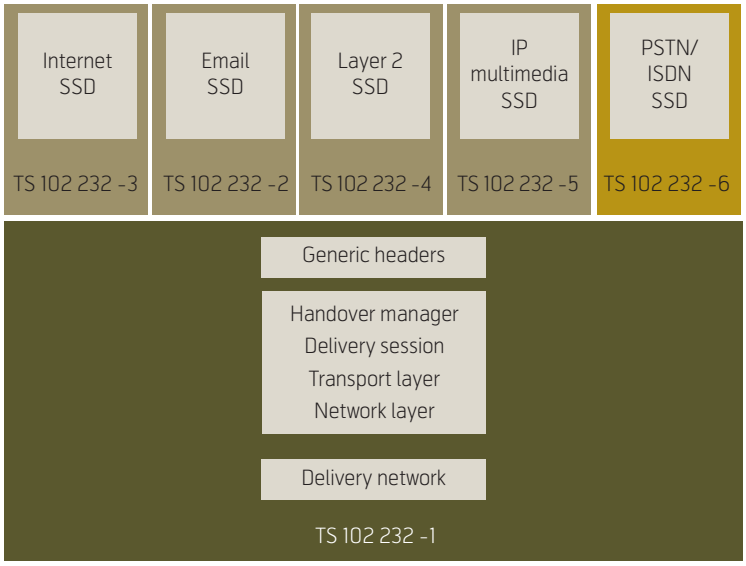


Figure 3 Lawful Interception specifications

Acknowledgements

The authors are grateful for permission to use material provided by Peter van der Arend (Chairman TC LI) and Dionisio Zumerle (ETSI LI Support Officer).

References on Lawful Interception

Published by TC LI (see www.etsi.org)

1 *Lawful Interception (LI); Handover Interface for the Lawful Interception of Telecommunications Traffic*. ES 201 671

2 *Lawful Interception (LI); Requirements for Network Functions*. ES 201 158

3 *Lawful Interception (LI); Handover Specification for IP Delivery*. TS 102 232

4 *Service-specific details for e-mail services*. TS 102 233

5 *Lawful Interception (LI); Service-specific details for internet access services*. TS 102 234

6 *Lawful Interception (LI); Handover interface for the lawful interception of telecommunications traffic*. TS 101 671

7 *Lawful Interception (LI); Requirements of Law Enforcement Agencies*. TS 101 331

8 *Lawful Interception (LI); Notes on ISDN lawful interception functionality*. TR 102 053

9 *Lawful Interception (LI); Issues on IP Interception*. TR 101 944

10 *Lawful Interception (LI); Concepts of Interception in a Generic Network Architecture*. TR 101 943

11 *Lawful Interception (LI); Service-specific details for Layer 2 Lawful Interception*. TS 102 815

12 *Lawful Interception (LI); ASN.1 tree structure of the Security Domain.* ETSI TR 102 503

Published by Other Technical Bodies

13 *Intelligent Networks (IN); Lawful Interception* [TC SPAN] EG 201 781

14 *Telecommunications and Internet Protocol Harmonisation Over Networks (TIPHON) Release 3; Service independent requirements definition; Lawful interception – top level requirements* [EP TIPHON] TR 101 772

15 *Telecommunications and Internet Protocol Harmonisation Over Networks (TIPHON™); Security; Studies into the Impact of lawful interception* [EP TIPHON] TR 101 750

16 *Terrestrial Trunked Radio (TETRA); Security; Lawful Interception (LI) interface* [EP TETRA] EN 301 040

17 *Terrestrial Trunked Radio (TETRA); Security; Lawful Interception (LI) interface; Feasibility study report* [EP TETRA] EG 201 040

18 *Digital cellular telecommunications system (Phase 2+); Lawful interception requirements for GSM* (GSM 01.33 version 8.0.0 Release 1999) [TC SMG] TR 101 514

19 *Digital cellular telecommunications system (Phase 2+); Lawful interception – Stage 1* (GSM 02.33 version 8.0.1 Release 1999) [TC SMG] TS 101 507

20 *Digital cellular telecommunications system (Phase 2+); Lawful Interception; Stage 2* (3GPP TS 43.033 version 5.0.0 Release 5) [3GPP SA3] TS 143 033

21 *Digital cellular telecommunications system (Phase 2+); Lawful Interception; Stage 1* (3GPP TS 42.033 version 5.0.0 Release 5) [3GPP SA3] TS 142 033

22 *Digital cellular telecommunications system (Phase 2+); Lawful Interception requirements for GSM* (3GPP TR 41.033 version 5.0.0 Release 5) [3GPP SA3] TR 141 033

23 *Universal Mobile Telecommunications System (UMTS); 3G security; Handover interface for Lawful Interception (LI)* (3GPP TS 33.108 version 5.4.0 Release 5) [3GPP SA3] TS 133 108

24 *Universal Mobile Telecommunications System (UMTS); 3G security; Lawful interception architecture and functions* (3GPP TS 33.107 version 5.5.0 Release 5) [3GPP SA3] TS 133 107

25 *Universal Mobile Telecommunications System (UMTS); Lawful interception requirements* (3GPP TS 33.106 version 5.1.0 Release 5) [3GPP SA3] TS 133 106

26 *Digital cellular telecommunications system (Phase 2+) (GSM); Lawful interception; Stage 2* (3GPP TS 03.33 version 8.1.0 Release 1999) [3GPP SA3] TS 101 509

27 *AT Digital; Digital Broadband Cable Access to the Public Telecommunications Network; IP Multimedia Time Critical Services; Part 20: Lawful Interception; Sub-part 1: CMS based Voice Telephony Services*. TS 101 909-20-1

28 *AT Digital; Digital Broadband Cable Access to the Public Telecommunications Network; IP Multimedia Time Critical Services; Part 20: Lawful Interception; Sub-part 2: Streamed multimedia services*. TS 101 909-20-2

Published by The European Union

29 European Council Resolution, January 1995, JAI 42 Rev 28197/2/95, published in the Official Journal reference 96C 329/01, 4 November 1996

Miscellaneous

30 *ETSI Lawful Interception homepage*: http://portal.etsi.org/li/Summary.asp

Rupert Thorogood (C.Eng., FIEE) is a Professional Telecommunications Engineer who worked in international telecommunications for 35 years with Cable and Wireless plc in many different countries. After retiring from Cable and Wireless he became a Consultant with the UK Home Office to assist in their work on lawful interception. He was a founder member of the ETSI Lawful Interception Technical Committee and was secretary for ten years.

email: rthorogood1@compuserve.com

Charles Brookson (C.Eng., FIEE FRSA) is Assistant Director in the Department of Trade and Industry and is a Professional Electronic Engineer. He previously was Head of Security for one2one (now T-Mobile UK), and worked within British Telecom for 20 years before. He has worked in many security areas over the last 30 years. Charles Brookson has been Chairman of the GSM Association Security Group. He has been working the GSM and 3GPP security standards, first chairing the Algorithm Expert Group in 1986. He is Chairman of the NISSG, a group that was set up to co-ordinate security standards amongst the three European Security Standards organisations and other bodies outside Europe. He is also Chairman of ETSI OCG Security, which is responsible for security within ETSI, and is on the Permanent Stakeholders group of ENISA, The European Network and Information Security Agency.

email: charles@zeata.co.uk

Privacy and the Regulatory Big Brothers

GEIR M. KØIEN

Geir M. Køien is a researcher in the Network Technologies group of Telenor R&I

They who can give up essential Liberty to obtain a little temporary Safety, deserve neither Liberty nor Safety [1].
- Benjamin Franklin

The Justification for Regulatory Control

Lawful Interception (LI), as a measure against organized crime and terrorist activities, is largely seen as a just measure and most people would agree that lawful interception is a necessary tool in the fight against organized crime etc. The threats against society from organized crime and terrorist activity are real and in the post 9/11 era most people will recognize that the law enforcement agencies must be given the proper tools to protect society against serious criminal activities. As described by Thorogood and Brookson in the article *Lawful Interception* [2], the European Lawful Interception regime has its foundation in the European Council Resolution of January 1995 [3], which outlined the requirements for lawful interception.

Many people will no doubt also defend the 'anti-terrorist' measures like the EU Data Retention Directive (DRD) [4]. Others are not equally convinced that wholesale surveillance of all communication can be justified on the off-chance that some of us may be terrorists. There are however quite a few practical problems with the approach in addition to the larger moral and ethical issues with this type of surveillance of the population. See also the articles by Svendsen [5] and Rognsvåg [6].

It is noted that the networks are able to determine the identity and position of a caller. This is not a problem per se as there are many useful services that require this information (so-called *location-based services*). Providing that one have positive user consent this is perfectly acceptable. There is one exception to the explicit 'user consent' rule and this is the emergency call (E112/E911) service. However, the E112/E911 service is a special case and the justification is a convincing one. Given the circumstances for an emergency call most people would agree that the emergency services should be granted access to subscriber identity and subscriber position information. An excellent starting point for information on emergency services is the ETSI EMTEL homepage [7].

The Dangers of Introducing 'Security Holes' in the Networks

Irrespective of the justification for lawful interception etc there are architectural problems with the introduction of 'security holes' in the system architecture.

One problem with designing interception points in the system architecture is obviously that it makes it easy for the law enforcement agencies to intercept communication. It has of course always been possible to intercept telecommunication, but with a fully digitalized system architecture the problem is that interception will scale. The interception methods that were used for old POTS[1)] systems did not scale very well at all as they required manual intervention. Technically, LI in a digital system can be made to scale well and there may be a temptation to broaden the scope. Particularly, one may fear that in repressive regimes the temptation to intercept communication from (non-criminal) political opponents may be too great to resist. Other uses would be state endorsed surveillance of foreigners and foreign businesses. There are regimes which are happy to participate in industrial espionage to protect local companies. There are of course gray areas here, but particularly in areas of state security and in military affairs there is a tendency to permit practices that is tangent to industrial espionage.

Too liberal use of LI for otherwise legitimate reasons is also problematic. LI represents a very serious privacy violation and even suspected criminals have rights. Most people would therefore agree that LI should only be used for serious crime and that use of LI for petty crime is not generally justified. There is of course also the issue of plausible case. How much circumstantial evidence of a serious crime is necessary before use of LI can be authorized? It can be hard to draw that line and what is generally deemed acceptable may differ substantially from the actual practice. The acceptance may vary over time and it is influenced by factors such as perceived threat level, media coverage and the general political climate.

1) *POTS is a common abbreviation for Plain Old Telephone Service.*

 ISSN 0085-7130

Needless to say, acceptance of LI usage will very from country to country too.

Of course, people working for telecom operators aren't necessarily all law abiding citizens either. There have been reports of corrupt telecom employees that have sold sensitive information about individuals and companies. Tellingly, the publicized cases have been about selling information about the location of celebrities to paparazzi photographers, etc. Obviously, information about the conversations of a CEO or a minister of state could be valuable to other parties too.

The justification for LI is nevertheless a solid one. Furthermore, it is a targeted and subscriber specific surveillance service. The DRD and similar schemes, on the other hand, are much more broad in scope and aim at the whole subscriber population. This is a 'guilty till proven guilty' approach. The amount of data that is being collected varies, but generally what is collected is control data and not the conversation as such. Still, this is a massive invasion of people's privacy and it takes place irrespective of whether you are a suspected criminal/terrorist or not. Now, terrorist activity is certainly a very serious crime and it represents a threat not only to individual safety, but to society itself so equally drastic measures may be necessary. The DRD is surely a draconian measure, but there is nevertheless quite a lot of popular support for these types of measures. The argument goes along the lines of 'if you are innocent then you have nothing to fear'. This is all fine, but again there are powerful counter arguments along the lines that innocent people have a right to privacy. The famous Franklin quote [1] illustrates that this problem has existed for some time. There is a real fear that powerful tools like the DRD, used with data mining tools and in combination with other databases etc, permit scenarios reminiscent of Orwell's 1984 [8] totalitarian surveillance state.

So the question remains; should we trust Big Brother? And, realistically, what is the alternative? How much privacy *must* we be ready to give up to obtain a minimum level of safety? The jury is still out, and of course there is no definitive answer, just difficult questions.

References

1 Franklin, B. *Files as "Contribution to Conference (III)"*, 17 February 1775. (It is noted that Franklin used similar phrases in other contributions.)

2 Thorogood, R, Brookson, C. Lawful Interception. *Telektronikk*, 103 (2), 33-36, 2007. (This issue)

3 European Council. Resolution January 1995, JAI 42 Rev 28197/2/95, published in the Official Journal reference 96C 329/01, 4 November 1996.

4 DRD European Council. Resolution January 1995, JAI 42 Rev 28197/2/95, published in the Official Journal reference 96C 329/01, 4 November 1996.

(This is a very useful homepage for anybody interested in emergency communication over public networks.)

5 Svendsen, B. The EU Directive on data retention – An end to justify the means. *Telektronikk*, 103 (2), 31-32, 2007. (This issue)

6 Rognsvåg, K. Privacy and protection of personal data. *Telektronikk*, 103 (2), 27-30, 2007. (This issue)

7 ETSI Emergency Communications homepage. http://www.emtel.etsi.org/

8 Orwell, G. *Nineteen-Eighty-Four*. Secker & Warburg, 1949. (Later edn. ISBN 0-451-52493-4, 1949)

For a presentation of Geir M. Køien, please turn to page 3.

Subscriber Privacy in Cellular Systems

GEIR M. KØIEN

Geir M. Køien is a researcher in the Network Technologies group of Telenor R&I

The 3G cellular access security architectures do not provide satisfactory subscriber privacy and fail to reflect the fact that there are three principal entities involved in the security context. In this article a beyond-3G Privacy Enhanced 3-Way Authentication and Key Agreement (PE3WAKA) protocol is presented. The PE3WAKA protocol provides substantially improved user privacy and a 3-way security context.

I Introduction

Motivation

In this article we analyze the standard 3G access security architecture. We then propose new requirements that strengthen the security and improve the subscriber privacy. To achieve enhanced subscriber privacy the subscriber identity presentation scheme has been modified. Satisfactory performance is a decisive factor in any real-world system, and the design of the new protocol combines selected Mobility Management signaling sequences with the Authentication and Key Agreement (AKA) protocol to achieve a fast setup. The PE3WAKA protocol presented here belongs to a family of protocols. The other family members are presented and discussed in [1-3].

The paper is structured as follows: Section II covers the security requirements, section III covers elements of the cellular environment, section IV is on the need for a security context hierarchy, section V is subscriber privacy, section VI is on the cryptographic basis for the PE3WAKA protocol and VII describes the PE3WAKA protocol. Section VIII contains an analysis of the protocol, and IX is the conclusion.

The reader is assumed to be reasonably acquainted with cellular system architectures.

Cellular Access Security

The Principal Entities

Access security is mainly concerned with ***a)*** *authentication and key agreement* between the User Entity (UE), the Home Entity (HE) and the Serving Network (SN), and ***b)*** *link-layer access protection* between the UE and the SN. The cellular model has three principal entity types. In a typical access signaling scenario exactly one instance of each entity type is involved.

- User Entity (UE)[1)]
  This is the mobile device, including a security module (typically a smartcard). The UE has a subscription with a Home Entity. The mobile device may be able to access multiple network types, including GSM/GPRS, UMTS and IEEE 802.11 based WLANs.

- Home Entity (HE)[2)]
  The HE manages the UE subscription data, the UE location data and UE service charging. The HE has security jurisdiction over the UE, and the HE assigns both the permanent UE identity and the corresponding security credentials.

- Serving Network (SN)
  The SN is the physical network that provides access for the UE. An SN will permit a UE to roam onto its network provided the HE and the SN have a roaming agreement. The SN handles local mobility management (i.e. *handover* and *location updating*). The SN entity is operated by an SN operator.

A cellular service provider commonly owns both SN and HE entities, but we assume that HE and SN in general are operated by separate administrative entities.

The Industry Standard

The 3G systems are the current industry standard for cellular systems. The main access security services provided are [4-7]:

- Mutual entity authentication
  The entity authentication is between the UE and the network(s). At the UE the actual entity is the USIM, which is the security- and subscription application running on the smartcard (UICC). At the network side the SN will execute the *challenge-response* mechanism after having received the

1) *The corresponding 3GPP term is* User Equipment.

2) *The corresponding 3GPP term is* Home Environment.

 ISSN 0085-7130

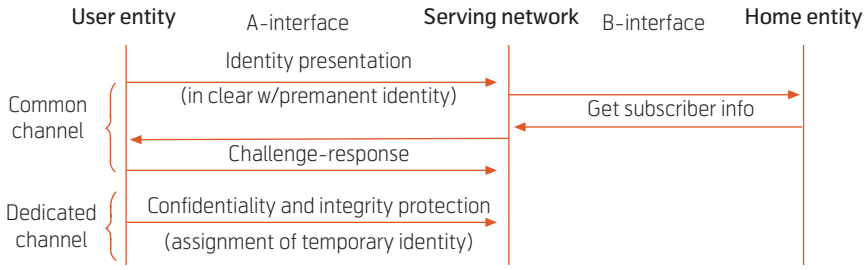


Figure 1 Overview of the 3G Authentication and Key Agreement Procedure

security credentials (called the *Authentication Vector (AV)* in 3G) from the HE. The HE is offline with respect to the challenge-response part of the 3G AKA procedure.

- Data Integrity and Data Confidentiality on the access link
  The 3G systems provide data confidentiality on (almost) all subscriber related data between the UE and the RNC node in the SN network. The 3G network also provides data integrity[3)] protection for the control plane messages. This is provided using symmetric-key encryption techniques.

- Location/Identity Privacy
  The 3GPP security architecture specification does include requirements for *subscriber identity privacy* and *subscriber location privacy* [4], but the actual subscriber privacy offered falls short of the requirements and must be considered inadequate [5,8].

There are important limits to both subscriber privacy and the access security in the 3G systems. A number of these shortcomings are directly related to the history behind the development of the 3G systems and the 2G legacy[4)]. In the PE3WAKA protocol example we address many of these weaknesses and provide enhanced subscriber privacy and effective and efficient key agreement between the principals.

II Security and Privacy Requirements

Location/Identity Privacy

In the 2G/3G systems the subscriber identity presentation is permitted to be in cleartext over the radio interface. In fact, the permanent identity (IMSI) is routinely exposed over the radio interface. Subsequent to the initial presentation the system (the SN) will assign a temporary identity (TMSI) to the UE. The TMSI is assigned subsequent to starting the security services. Thus, the TMSI is provided to the UE in ciphertext form. The TMSI is then to be used in cleartext in *paging* and *access request* procedures. This provides a measure of subscriber identity privacy against a passive intruder (plain eavesdropping), but the scheme does not provide protection against active attacks and it does not prevent the IMSI from exposures.

Furthermore, the 3G standards entirely fail to recognize the need for privacy protection from the cellular network entities. As documented in [8-10], the UE should not need to expose both its location and identity to the SN and the HE. We have identified a set of high-level requirements relating to subscriber privacy. The requirements are described below in terms of the actors in the environment.

- Intruder
  An intruder should never be allowed to learn the identity or the location of a UE. The protection must be effective against both passive and active attacks. One cannot entirely prevent an intruder from detecting the presence of a UE in an area and the intruder cannot entirely be prevented from tracking the physical (radio) presence of a UE. Short of that, the intruder should not be allowed to gain information from layer 2 and 3 that permits the intruder to infer the identity of the UE or to derive a tentative identification which may be used for tracking the un-identified user.

- HE
  The HE issues the permanent UE identity *(UEID)*. The HE will also know the public SN identity and it *may* (for routing purposes) know the SN server area where the UE is located.

  However, we postulate that the HE does not need to know the precise UE location. That is, the UE

3) *The data integrity protection includes message origin authentication and data integrity.*

4) *The 2G control model is very much oriented towards provision of circuit-switched services. Of course, the 2G systems, like GSM, were originally designed to be cellular equivalent to the ISDN system.*

should be permitted to retain a certain measure of location privacy from the HE.

- SN
  The SN will, due to radio signal measurements etc, necessarily know approximately where the UE is located whenever there is contact between the UE and the SN. Since there is no practical way for the subscriber to know when there is lower layer radio contact between the UE and the SN, the subscriber must assume that the SN will know its location. However, there is no compelling system architecture argument that requires the SN to know the permanent UE identity. The primary SN requirement is that the HE recognizes the UE and accepts liability (charging/billing) for the UE. Consequently, an anonymous UE alias identity will suffice to cater for the identification needs of the SN.

- Lawful Interception (LI)
  LI capabilities are a mandatory requirement for public cellular service operators. The LI requirements include mandatory provision of subscriber location information. It is additionally noted that other regulatory requirements may also apply, like the EU Data Retention Directive etc. The practical consequence of this is that the provided subscriber privacy must be revocable (this also applies to emergency calls).

Home Control, Online Authentication and Spatial Binding

Home Control

The 2G/3G AKA protocols are off-line with respect to the HE. Authentication is effectively delegated to the SN. In the early 2G days this could be defended, but today it is hard to justify the amount of trust required. From a HE perspective one therefore needs to improve the HE control. One of the most obvious ways to improve the situation is to require the AKA protocol to be an online protocol. The HE would then be in direct contact with the UE at each authentication event.

3-Way Security Context

There are three principal entities in the cellular model. The standard off-line 3G AKA does not capture this and there is no clear distinction between the HE and the SN [4,5] as seen from the UE point of view. The UE therefore only knows that it is in contact with an SN who has obtained a valid *Authentication Vector (AV)*. The AV was obtained at some point in time, but there is no 100 % certain way to determine if the HE still acknowledges the AV. That is, there is no true AV freshness guarantee or credible revocation mechanism in place.

In 3GPP one has developed a variant of the UMTS AKA protocol. This AKA protocol is used for the 3GPP-WLAN interworking case. The modified protocol is effectively an online protocol, and the problem seems to have been solved. However, this protocol also has its problem and in particular it leaves the SN with almost no authority [11]. The SN, which after all is the party that provides the access service, also has a legitimate need for control. To improve home control while retaining SN control, an online 3-way AKA protocol is required.

Spatial Context Binding

The security context should have limited validity. Protocols like IPsec [12] have exposure restrictions with respect to protection usage (KByte/packets) and lifetime (seconds). For a mobile user one can extend the exposure control to a spatial dimension. The spatial resolution should be of 'reasonable' granularity. The meaning of 'reasonable' is system dependent, and would be a trade-off between signaling performance/workload and privacy aspects. A useful compromise between exposure control and performance seems to be to assign the spatial binding to the SN server area (this is analogous to binding the context to an SGSN/VLR area in 3G). We note that Home Control requirements may be at odds with UE privacy requirements.

Performance

The security procedures must be designed to meet the overall system performance requirements. We are particularly concerned about the accumulated delays during initial registration and user session establishment. The 3G systems were developed as evolutions from an existing 2G base. The 2G system limitations that determined and justified the sequential 2G call set-up no longer exists. In our revised beyond-3G set-up scheme we therefore propose to combine selected mobility management procedures with the security setup. The net effect is that the total number of round-trips can be reduced, even when deploying an online privacy enhanced 3-way AKA protocol.

III The Cellular Environment

Principals and Trust

We have the following *security* trust-relationships:

- UE – HE
  This is a subscription based relationship. The subscription contract forms a legal basis for the trust relationship. Regulatory and legal requirements apply to the subscription contract. The HE assigns the permanent UE identity and the long-term security credentials. From a security perspective we have that the HE has jurisdiction over the UE.

- HE – SN
  This relationship is based on legally binding roaming agreements. We assume these relationships to be long-term and we assume limited mutual trust.

- SN – UE
  There exists no a priori *SN – UE* relationship[5)]. We assume transitive trust-relationships, and this permits (on-demand) establishment of the *SN – UE* relationship. The *SN – UE* trust relationship is limited in scope and time.

Trust with regard to privacy is a different matter. From the UE perspective, the HE and SN should by default only be semi-trusted and the PE3WAKA protocol should therefore aim at concealing UE privacy sensitive data from the HE and SN. The UE may later choose to reveal its location and/or identity to obtain (location) context specific services etc, but this would be another matter entirely.

Intruder Model

Dolev-Yao Intruder

We assume that the Dolev-Yao (DY) [13] intruder model applies. That is, we assume an intruder that can read all data on all channels and can delete, substitute, add, replay or modify any message at will. Only proper application of the cryptographic protection in the PE3WAKA protocol will prevent the DY intruder from breaking the protocol. It is noted that since we want the PE3WAKA protocol to provide enhanced subscriber privacy then correspondingly we must consider the DY Intruder to be successful if subscriber privacy sensitive data is revealed by the protocol. Thus, our DY Intruder is 'aware' of the privacy issues.

The Main Mobility Procedures

To simplify matters we only consider the most basic procedures. We have four main mobility procedures in the classic cellular control plane models (which apply to GSM, UMTS etc).

Registration

The cells broadcast the network identity and area code. The UE is responsible for registering with the network, and this is done when entering a new area. The UE must identify itself during the registration procedure.

Data to UE

The UE in *idle* state will listen for *paging* on the paging broadcast channel. The paging is in cleartext and the paged identity will be visible to all entities within the paging area (the location area). The UE will respond to the paging and access the system. It must then identity itself (possibly in cleartext).

Data from UE

This procedure is in principle identical to the *Data to UE* procedure except from the paging part.

Handover

This procedure consists of seamless switching between radio-channels during active transfer. This procedure is local and performed by the UE and SN. The SN has global knowledge of availability of neighbor access point capacity etc, and we assume SN controlled handover.

Identity in Clear

In 2G/3G systems the permanent subscriber identity can, depending on circumstances, be presented in clear for three of the four procedures. In our beyond-3G scheme we will ensure that the permanent identity is never exposed over-the-air or indeed transferred in cleartext over any other interface.

Interfaces

In a real-world cellular scenario there would be a number of system interfaces. In our model we only attempt to capture the direct interfaces between the security principals (Figure 2). The A-interface is the over-the-air interface between the UE and the SN. It covers common channels and dedicated channels (to be protected). The common channels are public and unprotected, and are used during access setup. During the setup phase the A-interface may be severely band-width restricted. The B-interface is a high capacity (fixed line) interface. We assume that the HE and SN have pre-established a secure connection over the B-interface.

Protocol Integration and Round-trip Efficiency

The 3G AKA protocol is a one-pass protocol for the successful case. This is achieved at the cost of a sequence number replay-protection mechanism [4]. However, the apparent efficiency is an illusion when the whole set-up signaling is studied [8]. The 3G systems inherited a circuit-switched ISDN based service model designed for the 2G systems. The 2G control model had to take into account the limitations of the 2G radio-systems and SS7-based fixed-network signaling. Taken together this has led to a highly sequential set-up. One first completes the Radio-Resource Management (RRM) signaling to establish

5) *This is true for the generic case, but of course the HE operator may also have a home SN network. This home SN network may have a special standing, but with respect to execution of the control protocols there is nothing inherently special with a home SN network.*

the physical link. Then the Mobility Management (MM) procedures are run and finally the Cellular Access Security (CAS) services are run. The number of signaling round-trips to complete a set-up is therefore quite high in 2G/3G systems. This is not satisfactory for a beyond-3G system with a service model dominated by connectionless packet-switched services.

To remedy the situation one must re-design the control signaling. As pointed out in [14] there are lower bounds on the necessary number of round-trips for security protocols, but if one combines logically connected sequences one may be able to reduce the total number of round-trips. We shall not propose a new control model here, but will argue that several of the RRM, MM and CAS procedures will/can be triggered by the same physical/external events. Dependencies not withstanding, one can integrate some of the procedures and reduce the total number of round-trips. We shall focus on the MM Initial Registration procedure and the Authentication and Key Agreement procedures. In order to get access the UE must register with the SN (this includes identity presentation) and authenticate itself. The combined procedure must therefore provide identity presentation, registration, and establishment of the security context. It is noted that re-keying events may coincide with RRM events, but we will not investigate this topic further here.

IV Security Context Hierarchy

We classify the security context relations according to spatial or temporal coverage. In practice the two views yield similar structures. The three basic levels are outlined in Figure 3.

The Long-Term Context

There are two long-term contexts.

- HE-SN
  This context is based on roaming agreements and will be long-lived. The spatial validity of the *HE-SN* long-term context will generally be for the full coverage of the SN network.

- HE-UE
  This context is defined by the UE subscription. The context contains long-term security credentials. Spatially, the context is valid for all HE provided areas, which in principle is the sum of all HE roaming agreement areas.

The HE will have a set of HE-SN contexts (typically >100). The SN may likewise have agreements with a large number of HE operators. The HE may serve millions of subscribers (UEs). The UE will have exactly one HE association.

A-interface: Over-the-air-interface, w/limited capacity
B-interface: High-capacity authenticated and protected channel

Figure 2 Interfaces

Medium-Term Context

The medium-term context is dynamically established on the basis of the long-term contexts, and it is defined by the (UE,SN,HE)-tuple. The context is established in conjunction with the UE registration. The validity of the context is confined by the spatial validity (an Area Code (AC), related to registration area), a validity period (VP) and possibly a usage count (Kbytes/packets). The geographical coverage of AC would depend on the network topology. In our scheme, establishment of the medium-term security context is aligned with the initial registration procedure. We advocate that the area should be aligned with an aggregate server area (analogous to VLR/SGSN areas in 3G). The AC may consist of a set of location/paging areas. Subsequent *MM location updating* procedures may then be executed without re-establishing the medium-term context. The AC should be published on an SN broadcast channel. The validity period, *VP*, should be long enough to avoid excessive context invalidations yet not be substantially longer than normal UE presence in an area. The actual mobility patterns may vary over time and it should be possible to configure the *VP* parameter. We advocate that the SN determines the validity period parameter (VP_{SN}) and publishes it on a broadcast channel. The UE should additionally have a HE determined maximum (VP_{HE}) value, and the context should be canceled when the validity period expires ($VP = \mathrm{Min}(VP_{HE}, VP_{SN})$).

Short-Term Context

The short-term context only applies to the A-interface. It consists of symmetric-key key material for protection of user related communication (including

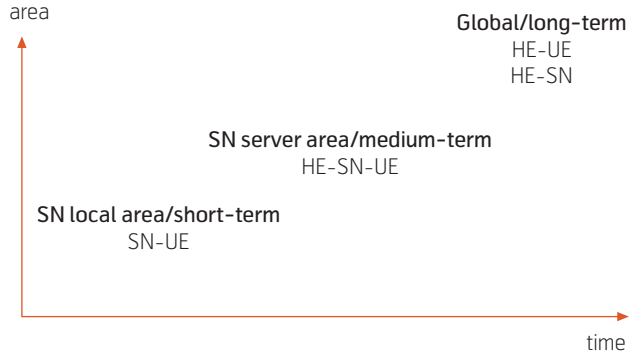


Figure 3 Security Context Hierarchy

ISSN 0085-7130 ©Telenor ASA 2007

user related control signaling) and an identity binding. The short-term context credentials are derived from the medium-term context. The life-time of this context may exceed the scope of atomic transactions, but it should nevertheless be short lived. The spatial scope should also be limited. Network topology considerations suggest that the geographical scope may be limited to an access point controller area, but the scope could also be aligned directly with the medium-term context.

V Privacy Matters

Privacy and Identity Presentation

Integration of the *MM Initial Registration* and *Authentication and Key Agreement* procedures will allow us to save round-trips compared to the 3G case. Given that the registration is the sole responsibility of the UE, we have that the combined procedure must be initiated by the UE. The UE must therefore be the initiator of the PE3WAKA protocol. This is in contrast to the 2G/3G AKA protocols, where the UE is always the responder.

The Medium-Term Context Identity

To avoid having the UE presenting itself with the permanent identity (*UEID*), we propose to let the UE identify itself with an anonymous and temporary Context Identity (*CID*). The *CID* value should be constructed using a pseudo-random function (*prf*), and for a given UE a particular *CID* value should only be used once. The *CID* will therefore only be used for exactly one (UE,SN,HE) medium-term context. To let the UE choose the *CID* means that the SN may experience *CID* collisions. The *CID* collision frequency experienced by a SN server should be very low.

The *CID* value space must therefore be correspondingly large. With no bias to the *CID* choices, we can use the approximation $p = k/m$, where p is collision probability, k the maximum number of subscribers within the SN server area and m is the space of the CID value. To be on the safe side, we assume that an SN server can serve one billion simultaneous users ($k = 10^9$). We do not anticipate frequent 3-way authentications, but even with the extremely unlikely rate of one AKA event every second and a system lifetime of 30 years there will less than one billion events (10^9)[6]. So if we require a collision to occur at most once in the lifetime of a subscriber we have $p = 1/10^9$. To be on the safe side the *CID* must then have a range of $m = 10^9 \cdot 10^9$. Given this, the *CID* information element (IE) can be stored in a variable with a minimum of 60 bits ($2^{60} \geq 10^{18}$). To allow some safety margin and to align with byte boundaries one may in practice decide on 64 bit as the minimum size for the *CID* IE.

$$prf(\cdot) \rightarrow CID \tag{1}$$

Privacy and the Context Identity

We want to conceal the permanent UE identity (*UEID*) from both the SN and external intruders. The UE must therefore be able to provide the (*UEID,CID*) association to HE while preventing the *UEID* from being exposed. Additionally, the *CID* should not be exposed over the interfaces. To solve the problems the UE must be able to privately communicate *UEID* and *CID* to the HE. The SN must be informed about the *CID*, and furthermore the HE must corroborate to the SN that *CID* is a valid UE identity.

Anonymous Tracking

If an identity is used for a prolonged period an intruder may be able to track the user. The use of pseudo-anonymous identities will prevent the external intruder from knowing the permanent UE identity, but the intruder may still find reason to track the anonymous user. To protect the UE against anonymous tracking we require that the UE and SN use a temporary alias identity (*AID*) for cleartext presentation. The SN assigns the *AID* in confidentiality protected form. The *AID* will be used, in cleartext, for paging- and access request purposes. The *AID* should ideally be assigned for one-time use, but it may be used a limited number of times before being replaced by a new *AID*. There should be no apparent correlation between the (*UEID,CID*)-tuple and the (sequence of) *AID*(s), and there should be no apparent correlation between any *AID* in the sequence of *AID*s.

Temporary Alias Identity (*AID*):

$$prf(\cdot) \rightarrow AID \tag{2}$$

VI Cryptographic Basis

Use of Public-key Cryptography

In order for the UE to privately communicate its permanent identity (*UEID*) to the HE the UE must construct a message with *UEID* in ciphertext. This represents a problem in the sense that the HE must be able to decipher the message without knowing who the message originator is. The problem can be solved with symmetric group keys or with asymmetric cryptography.

6) *Assuming a year with 365.25 days, the number of seconds is: $60 \cdot 60 \cdot 24 \cdot 365.25 \cdot 30 \approx 947$ million (seconds).*

Symmetric-key group keys are messy in terms of management etc. It also means that all UEs must share the same key. This is clearly contrary to our privacy requirements since it would allow all group-key holders to decipher the message.

The proposed solution is therefore to let the UE use (asymmetric) public-key encryption. The encryption will be with the public part of the public-key (*HEpub*) belonging to the HE. Only the HE has access to the corresponding private key (*HEpri*). Thus, only the HE is able to decipher the message. Observe that any entity knowing the HE public key (*HEpub*) will able to construct an encrypted message. The HE cannot therefore assume that the encrypted message comes from the claimed UE. The message contents must therefore additionally have independent data integrity protection. This can be achieved by application of a *Message Authentication Code* (*MAC*) function. The MAC function computes a cryptographic checksum under the control of a secret key. The secret key is then a shared key between the UE and the HE.

Key Agreement

The Diffie-Hellman Key Exchange

It is suggested to use a Diffie-Hellman (DH) key exchange to produce the shared secret that is the basis for the medium-term security context. The DH procedure [15] is based on public-key cryptography.

The DH public keys needed to produce a shared secret can become large. We want to use the DH-shared secret (*dhs*) as the basis for producing session key material (data confidentiality key and data integrity key). We do not require perfect-forward secrecy (PFS) for the short-term contexts within the scope of a medium-term security context. We therefore allow re-use of the *dhs* for generation of short-term contexts between the UE and the SN during the lifetime of the medium-term context. To permit this we require that the *dhs* have at least 256 significant bits. This will amount to exchange of DH public keys of approximately 15K bit in size.

Unconventional Use of the DH Exchange

The common channels available during the early setup phase are likely to be severely bandwidth restricted. We therefore cannot expect to be able to carry out the DH exchange with 15K bit public keys over the A-interface. It is possible to save bandwidth by using ECC-DH (which requires only approximately 600 bits per key [16]), but we still need to conserve bandwidth over the A-interface. Given that HE has security jurisdiction over the UE, we may allow the HE to carry out the DH exchange on behalf of the UE. The HE must then communicate the *dhs*, in confidentiality protected form, to the UE after the DH exchange has taken place. We denote the public-key HE parameter DH_{HE} and the SN parameter DH_{SN}. DH group agreement etc should be part of the HE-SN long-term security context.

Symmetric Key Agreement

The symmetric key agreement scheme will use the DH secret (*dhs*), an identity and an area code as the basis. For the medium-term context the area code is the *AC*. For the short-term context a local area code (*LAC*) is used. This area code may be *the location/paging area* code.

Derivation of medium-term context and short-term (session) context symmetric keys:

$$\text{UE,SN: } KeyDerive_{dhs}(CID,AC) \rightarrow mtk \quad (3)$$

$$\text{UE,SN: } KeyDerive_{dhs}(CID,AID,LAC) \rightarrow stk \quad (4)$$

The function *KeyDerive* is a one-way key derivation function. The *mtk* (medium-term key) is only used between the UE and SN during the execution of PE3WAKA. The *stk* keyset and the *AID* are valid for the duration of the short-term context. For sake of simplicity we have presented *stk* and *mtk* as single keys, but in reality it would denote a key-set including both confidentiality- and integrity keys; the keys may additionally be directional ($UE \rightarrow SN$ and $SN \rightarrow UE$).

Authentication

Challenge-Response

To avoid excessive use of the asymmetric methods we let the UE and HE share a secret, *KA*, to be used as the basis for the two-way checksum-based challenge-response mechanism. The shared secret *KA* is also used for computing cryptographic checksums. Another HE-UE shared keyset, *KC*, is used for data confidentiality.

Construction of challenge data (CH_{UE},CH_{HE}):

$$\text{UE: } prf(\cdot) \rightarrow CH_{UE} \quad (5)$$

$$\text{HE: } prf(\cdot) \rightarrow CH_{HE} \quad (6)$$

The output of the pseudo-random function must be unpredictable and never repeat for the HE-UE context. The response is computed under the control of the authentication key (*KA*) and includes the *CID* in the input. The inclusion of *CID* is necessary to ensure proper binding between the medium-term context and entity authentication.

Computation of response data (RES_{UE},RES_{HE}):

$$\text{UE,HE: } \text{Resp}_{\text{KA}}(CID, CH_{UB}) \rightarrow RES_{UB} \quad (7)$$

$$\text{UE,HE: } \text{Resp}_{\text{KA}}(CID, CH_{HB}) \rightarrow RES_{HB} \quad (8)$$

The function *Resp()* is a keyed one-way function. The key *KA* is included in the long-term security context between the UE and HE.

HE-SN Authentication

We assume that secure communication over the B-interface has been established prior to PE3WAKA execution and that the SN and HE are mutually authenticated.

SN-UE Authentication

To protect its interests the SN must insist that the UE be authenticated. We shall allow this to mean that the SN has assurance that the HE accepts the *CID* identity as a valid UE identity. The SN already knows that HE has jurisdiction over the UE, and if the HE has corroboration of the UE identity then SN shall accept *CID* as a valid UE identity. This information is conveyed to the SN over the B-interface. The SN and HE have mutually created the DH secret *dhs* online, and they both believe that *dhs* is a fresh shared-secret for the *CID* medium-term context. If the UE, which claims the *CID* identity, can show proof of possession of *dhs*, then the SN shall be compelled to believe that the UE is the principal entity that HE has assigned the *CID* identity to.

Likewise, the UE knows that the *CID* is fresh. Given that the UE has received the *dhs* from the HE in encrypted form, the UE also believes that *dhs* is a valid shared secret for the *CID* context. Given the trust relationships and the HE jurisdiction over the UE, the UE will accept the SN (*SNID*) as an authenticated entity provided it can show proof of possession of the *dhs* and knowledge of *CID*. Proof of possession of *dhs* will be demonstrated indirectly by the use of symmetric keys (*mtk*) derived under control of *dhs*.

VII An Example PE3WAKA Protocol

Main Objectives

The main objectives of the PE3WAKA protocol are as follows:

- Entity Authentication
  The protocol must provide (direct or indirect) mutual entity authentication between the (UE-HE) and (UE-SN).
- Key Agreement
  The protocol must be able to produce session key material for the protection of communication over the A-interface.
- External-party Subscriber Privacy
  The protocol must prevent an external intruder from learning the permanent UE identity and/or location and from tracking an anonymous UE.
- Internal-party Subscriber Privacy
  The protocol must prevent the SN from learning the permanent UE identity. (The SN will necessarily know the approximate UE position.)

  The protocol must not allow the HE to learn the precise UE position. (The HE will know the permanent UE identity.)

We assume that the HE and the SN do not collaborate to deceive the UE (although this may be a realistic scenario for the 'home network' case; then the SN and HE will be controlled by the same administrative entity).

Compliance with Lawful Interception (LI) Requirements

We assume that future public cellular systems must comply with LI requirements. It must therefore be possible to correlate the HE and SN information to permit a Law Enforcing Agency (LEA) to determine both the identity and position of any given UE. The PE3WAKA protocol permits the *UEID* and position data to be revealed given authorized HE and SN cooperation.

Outline of PE3WAKA protocol

Protocol Message Exchange

We now examine the PE3WAKA protocol in more detail. An overview is given in Figure 4.

The PE3WAKA protocol

1. Message M1

UE computes the validity period (VP), retrieves the HE public key (*HEpub*) and constructs the context identity (*CID*). UE then generates the challenge-response data, CH_{UE} and RES_{UE}. The checksum, CK_{M1}, is computed over the cleartext contents of *A*, the HE and SN public identities (*HEID* and *SNID*), validity period *VP* and the HE public-key keyset identifier, *PKID*.

$A := \{\text{UEID,CID,CH}_{\text{UE}}\}_{\text{HEpub}}$
$\text{CK}_{\text{M1}} := \text{MAC}_{\text{KA}}(\text{UEID,SNID,HEID,CID,CH}_{\text{UE}}\text{,VP,PKID})$

$\text{UE} \rightarrow \text{SN: M1(A,HEID,PKID,VP,CK}_{\text{M1}}\text{)}$

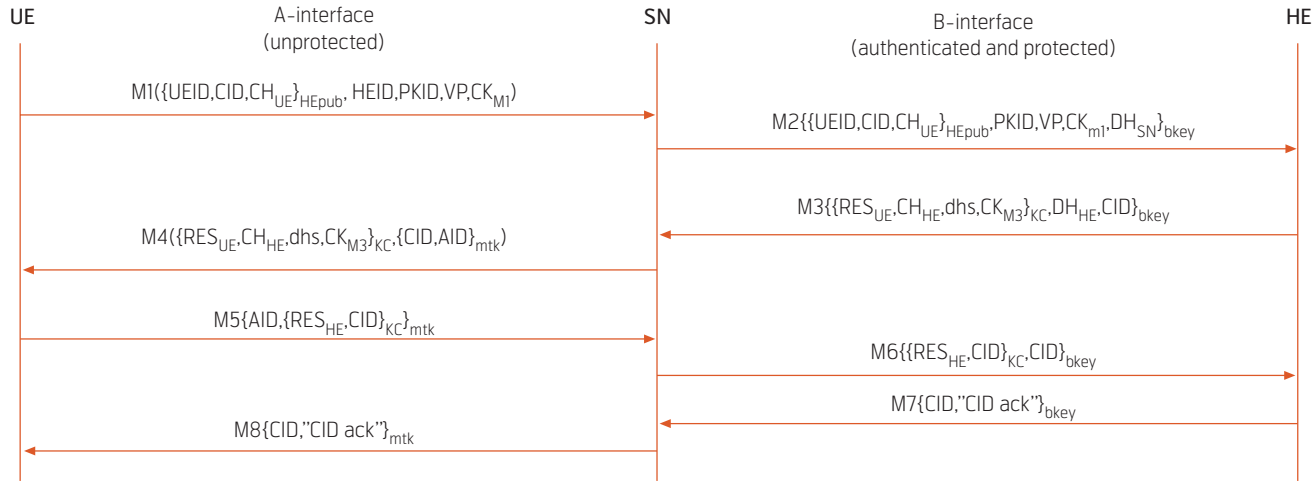


Figure 4 Outline of the PE3WAKA protocol

2. Message M2

The *VP* is checked for validity. HE address is derived from *HEID*. SN constructs DH key (DH_{SN}). The area codes (*AC* and *LAC*) are retrieved from the radio system.

$$SN \rightarrow HE: M2\{A,PKID,VP,CK_{M1},DH_{SN}\}_{bkey}$$

3. Message M3

HE retrieves the *HEpri* key identified by *PKID*. HE then decrypts *A*. HE now sees the claimed UE identity (*UEID*) and the associated context identity (*CID*). HE verifies the checksum CK_{M1}, before proceeding to compute the response, RES_{UE}, to the challenge, CH_{UE}. HE also computes a challenge to UE, CH_{HE} and a corresponding response, RES_{HE}. HE generates DH key (DH_{HE}) and computes the shared secret (*dhs*). The checksum CK_{M3} is used to provide *CID* binding and message origin authentication.

$$CK_{M3} := MAC_{KA}(RES_{UE},CH_{HE},dhs,CID)$$
$$B := \{RES_{UE},CH_{HE},dhs,CK_{M3}\}_{KC}$$

$$HE \rightarrow SN: M3\{B,DH_{HE},CID\}_{bkey}$$

4. Message M4

SN decrypts **M3** to get *B*, DH_{HE} and *CID*. SN computes *dhs* and generates the alias identity, *AID*. SN derives medium-term key *mtk* and short-term key *stk*.

$$SN \rightarrow UE: M4(B,\{CID,AID\}_{mtk})$$

5. Message M5

UE decrypts *B* and verifies that HE has confirmed *CID* as the context identity. UE verifies the response, RES_{UE}, and computes the response, RES_{HE}, to the challenge, CH_{HE}. Based on *dhs*, *CID* and *AC* the UE generates *mtk*, which it uses to decrypt the reminder of message **M4**. The UE now *sees* the *AID* and verifies that it is bound to *CID*. With *dhs*, *AID* and *LAC*, the UE generates session keys, *stk*.

$$UE \rightarrow SN: M5\{AID,\{RES_{HE},CID\}_{KC}\}_{mtk}$$

6. Message M6

SN decrypts **M5** and verifies *AID* assignment. By now the UE-SN short-term context is available. The SN still needs the final confirmation from HE that *CID* is authenticated, but the SN itself essentially has this confirmation given that UE has demonstrated possession of *dhs* (which the SN knows to be fresh). The short-term context can therefore tentatively be activated.

$$SN \rightarrow HE: M6\{\{RES_{HE},CID\}_{KC},CID\}_{bkey}$$

7. Message M7

HE decrypts **M6** and decrypts the UE response. HE then verifies the response, RES_{HE}.

$$HE \rightarrow SN: M7\{CID,"CID\ ack"\}_{bkey}$$

8. Message 8

SN verifies the HE acknowledge. Message M8 need not be a standalone message, but may be piggybacked on other messages.

It is noted that proof-of-possession for the *stk* keyset is not included. This is however achieved as soon as the session keys are used.

$$SN \rightarrow UE: M8\{CID,"CID\ ack"\}_{mtk}$$

VIII Analysis of the PE3WAKA protocol

Authentication and Identity Handling

The 3G AKA protocol is a one round-trip protocol. This apparent efficiency is possible since identity presentation and the acknowledge messages are not directly part of the protocol. Furthermore, the 3G AKA protocol must rely on a sequence number scheme for replay protection. The use of a sequence number scheme is not unproblematic. It adds complexity and provides weaker replay protection [15]. We therefore decided to avoid the use of sequence numbers and related schemes. The mutual entity authentication taking place between the UE and the HE is based on standard double challenge-response methods. The construct used is straightforward and we expect the method to be sound. The challenge-response method is based on keyed checksums, where the (UE-HE) shares an authentication key (*KA*). The inclusion of *CID* provides context binding and assists in assurance of freshness. The HE and UE are therefore compelled to believe in the authentication and to accept *CID* as a valid context identity.

The authentication between the UE and SN is novel and complex. The identity corroboration is indirect and relies on the transitivity of the SN-HE trust relationship and the fact that HE has jurisdiction over the UE. Inspired by the BAN notation [17] we informally have: SN participates in generation of the DH shared secret (*dhs*). The *dhs* is directly associated with the context identity *CID*. The PE3WAKA protocol is an online protocol and the B-interface is authenticated and protected. The SN therefore has reason to believe that the *dhs* is a fresh shared secret. The SN believes that HE has jurisdiction over the UE. This includes a belief that HE can communicate securely and privately with the UE. When the SN receives message M5 it has proof that (the presence of alias identity *AID*) UE possess the *dhs*. The SN therefore accepts that *CID* is an authenticated identity and it believes that the HE will accept liability for the entity presenting itself with *CID*.

Key Derivation

Although we require a three way authentication scheme, we only really need PE3WAKA derived keys between two of the principals. The use of a suitable one-way function, taking the shared secret (*dhs*) as the key basis, and using the context identity (*CID*) and an area code as inputs should provide a good basis for a key derivation scheme. We are therefore confident that the method used for the production of session key material is sound. Subsequent local key exchanges would use the same mechanism as shown in the PE3WAKA protocol (*mtk* and *dhs* are available for this purpose).

Spatio-Temporal Context Binding

The medium-term security context is confined spatially by the area code (*AC)* and temporally by the validity period (*VP*). Both the SN and UE must verify that the context binding is valid before using the context. The short-term context is tied to the alias identity (*AID*) and to a local area code (*LAC*). A session may exceed the validity (spatially and/or temporally) of a short-term context. Re-keying is then necessary, and in our scheme this is connected to assignment of the alias identity (*AID*). Local re-keying would then commence as a consequence of *AID* assignment.
A session may also extend beyond the validity of a medium-term context. The PE3WAKA protocol must then be re-run to reestablish the context or one may allow the session to conclude (using a suitable grace extension) before terminating the context. The medium-term temporal context expiry is also observed by the HE. Subsequent to a *VP* expiry, the HE will no longer route data towards the UE unless it re-registers.

Location/Identity Privacy

The stated goal was to prevent an outsider (intruder) from learning the *UEID* through eavesdropping or manipulation of over-the-air data. The *UEID* is never transmitted in clear. Provided that the crypto-primitives are safe and secure, and that the intruder cannot gain access to the HE private key (*HEpri*), we shall consider this requirement to be fulfilled. We also required that the SN never learn the permanent UE identity. We observe that the *UEID* is confidentiality protected and that the private key is not available to the SN. The SN will therefore never learn the *UEID*.

Furthermore, the HE should not learn the UE location. The HE will necessarily know in which roaming network the UE is located. The spatial binding of the area code (*AC*) to the medium-term context could have been problematic, but it is only the keysets used for the A-interface that have a spatial binding. It is also noted that the spatial binding to the keysets are one-way; that is, knowledge of the keys does not reveal the area codes.

Prevention of subscriber tracking was also a privacy goal. We observe that from an information theoretic perspective one must prove the absence of any kind of correlation between each access (including usage patterns etc) to provide full prevention against tracking. This is outside the scope of the PE3WAKA protocol. Instead, we consider this goal to be satisfied provided that the short-term sessions are sufficiently short, i.e. that new alias identities (*AID*) are used for separate paging/access requests. Strictly speaking, the lifetime of active sessions must also be contained to completely fulfill this goal.

Communication Aspects

The A-Interface

A concern for the PE3WAKA protocol has been the capacity of the (radio) common channels of the A-interface. The information element bit-sizes given below are only an estimate:

- M1: The M1 message contains one public-key encrypted block (*A*) and the *HEID*, *VP*, *PKID* and CK_{M1} bit-fields. The size of the public-key cipher output may vary, but we shall assume it to be at least 1 kilobit. The identity *HEID* is expected to be 128 bit wide, the validity period *VP* may also consume 128 bit. The *PKID* can be quite short and we assume that it is encoded in a 32 bit-field. The checksum may be quite large, but in line with [18,19] we contend that 64-128 bit is sufficient for the intended purpose.

- M4: The M4 message consists of return data from HE (*B*), encrypted with symmetric methods (we assume 128 bit blocksize. The secret *dhs* would require 256 bit of storage and the checksum should require no more than 128 bit. If we assume that RES_{HE} and *CID* each require 64 bit storage, we have less than 600 bit for block *B*. The remaining data from SN should fit in 256 bits of storage.

- M5: This message contains response data to the HE and the *AID*. It should require no more than 512 bits of storage space.

As indicated above, the messages all require less than 2 kilobit. Even systems with severe bandwidth restrictions during the set-up phase should be able to accommodate the modest requirements of the PE3WAKA protocol.

The B-Interface

The B-interface is presumed to be a fixed line interface. We do not foresee any capacity problems here.

Computational Aspects

Instantaneous Demand

We assume that the DH group parameters are agreed a priori between HE and SN (the HE and SN may of course occasionally change the parameters). The SN and HE may therefore precompute the DH public keys and store them for usage when a registration event happens. We expect that the use of symmetric methods (encryption/decryption, key derivation) does not impose a high workload. Consequently, we do not worry about the load generated by symmetric methods during PE3WAKA execution. The instantaneous demand can therefore be reduced to:

1 UE
One public-key encrypt operations (M1).
The UE initiates M1, and the real-time requirements here are tied to radio environment conditions and user perceived performance.

2 SN
The SN must compute one DH parameter (M2).
The SN must compute the DH shared secret (M3) before it can progress with M4.

3 HE
The HE must execute one public-key decrypt operation (M2) before it can progress with M3. HE must also compute a DH parameter and the DH shared secret before transmitting M3.

The HE node is a powerful entity and it can easily carry out the required computations. The SN node computations are relatively modest and should be unproblematic. Still, the HE and SN nodes must support a high number of UEs. It may therefore be beneficial to have specialized hardware to execute the public-key computations. Overall, we do not see that the HE and/or SN should have any problems with the computational burden imposed by the PE3WAKA protocol. The UE is a relatively powerful computational platform these days. The UE, moreover, can also dedicate almost all of its resources to the setup procedure. Today, smartcards [16] are able to execute asymmetric primitives with relative ease. We therefore do not foresee any problems with having the UE carry out public-key operations.

Denial-of-Service (DoS)

The PE3WAKA protocol does not provide explicit DoS protection. The PE3WAKA protocol operates in an environment where it is easy for an intruder to carry out access-denial attacks simply by disrupting the radio transmission. Access-denial attacks are local in nature and since they typically do not scale we have deliberately not tried to avert this type of attack in the PE3WAKA protocol. To limit computational DoS attacks we suggest that the SN restricts the arrival rate of PE3WAKA invocations per access point. The HE, likewise, may limit the number of simultaneous PE3WAKA sessions from any given SN. Together, these measures should effectively prevent a computational DoS attack from scaling.

Initiator-Responder Resilience

The PE3WAKA protocol does not specifically aim at providing initiator or responder resilience [20], but informal analysis suggests that it is difficult for the principals to gain any advantage here. The UE does not influence the *dhs*, but it selects the *CID*. The *CID* is used for key derivation and thus the UE has an

influence on (all) the session keys. The SN does not know the *CID* when it generates the DH_{SN} parameter, so its influence is limited. The HE will know both *CID* and DH_{SN}, and it may potentially tailor a DH_{HE} to create a specific *dhs*. However, to control the session keys (*stk*) one must also know *AID*, and HE does not know *AID*. So the HE cannot control the session keys *stk*.

Round-trips

Our claim was that the PE3WAKA protocol be at least as efficient as the 3G scheme. This amounts to assessing the cost of a 3G location updating sequence including 3G AKA and comparing it to the PE3WAKA round-trip cost (ref. 3G TS 23.108 (ch.7.3.1) [21]):

- UE → SN: Location Updating Request
- SN → UE: Authentication Request
- UE → SN: Authentication Response
- SN → UE: Location Updating Accept

The 3GPP TS 23.108 specification only covers the UE-SN communication. The SN-HE part can be found in 3G TS 29.002 [22], and it constitutes two separate request-reply sequences where the SN first fetches the subscriber information and then the security credentials. In PE3WAKA, the subscriber information can be forwarded in parallel with message M3.

The PE3WAKA protocol performs slightly better than the 3GPP scheme on the UE-SN interface. Note that message M4 serves as location registration confirmation. On the SN-HE interface, the PE3WAKA protocol requires two passes. The 3GPP scheme also requires two passes, but they *may* be executed in parallel. We note that the UE-SN context is potentially operative after SN reception of message M5. The SN still wants HE confirmation, but it now has indirect *CID* confirmation. It is therefore safe for the SN to activate the short-term context. Consequently, we can defend the claim that PE3WAKA is at least as efficient as the comparable 3GPP sequences.

Related Research

An alternative scheme using group pseudonyms is investigated in [23]. This scheme is GSM specific and we do not consider it to be relevant for a beyond-3G protocol. In [24] the authors investigate the use of MIXes. However, MIXes typically have indeterministic delays and one would need a set of MIXes to achieve sufficient privacy.

Most of our privacy requirements are also captured in [9,25]. However, the solutions presented do not offer sufficient home control. When we compare with our scheme we find that our security context hierarchy is more flexible and will afford better home control. Our scheme is also very different from [9] when it comes to identity management. Another paper that deals with user privacy and wireless authentication is [26]. Again, we find that many of the requirements are similar to our requirements. The solutions proposed are quite different from our approach. A direct comparison between our PE3WAKA and the suggested protocols in [26] would be unfair since they deliberately only considered low-cost solutions while we aim at a complete solution with a redesigned identity scheme.

Finally, many of the requirements we have arrived at are also found in [10]. The solution suggested in [10] does not take into account the possibility of integrating security setup with MM procedures, and it has a very different approach to identity management.

IX Summary and Conclusion

We have presented and analyzed the requirements for a privacy enhanced authentication and key agreement protocol for use in beyond-3G public cellular systems. The proposed PE3WAKA protocol is capable of substantially improved user location/identity privacy compared to the 3G scheme.

The PE3WAKA protocol provides enhanced privacy from eavesdropping and manipulation by an outsider adversary and it provides a measure of user location/identity privacy with respect to the serving network and the home operator. To achieve this, the user identity presentation scheme and the initial registration procedures had to be modified from the scheme used in 2G/3G systems. However, this change also allowed performance improvements by integration of signaling procedures that is anyway triggered by the same physical events.

References

1 Køien, G M, Oleshchuk, V A. Location Privacy for Cellular Systems; Analysis and Solution. In: *Proc. of Privacy Enhancing Technologies workshop 2005*, Cavtat, Croatia, 48-58. Springer, LNCS 3856, 2006. (ISBN978-3-540-34745-3)

2 Køien, G M. Privacy Enhanced Cellular Access Security. In: *Proceedings of the 4th ACM workshop on Wireless security (WiSe 2005)*, Germany, Cologne, 57-66, ACM Press, 2005. (ISBN 1-59593-142-2)

3 Køien, G M. Privacy Enhanced Mobile Authentication. *Wireless Personal Communications journal*, 40 (3), 443-455, 2007.

4 3GPP. *3G Security; Security architecture*. Sophia Antipolis, France, 2006. (3G TS 33.102)

5 Køien, G M. An Introduction to Access Security in UMTS. *IEEE Wireless Communications Magazine*, 11 (1), 8-18, 2004.

6 Rose, G, Køien, G M. Access Security in CDMA2000, Including a Comparison with UMTS Access Security. *IEEE Wireless Communications Magazine*, 11 (1), 19-25, 2004.

7 Nyberg, K, Niemi, V. *UMTS Security*. Wiley, 2003. (ISBN 0-470-84794-8)

8 Køien, G M. Principles for Cellular Access Security. *NORDSEC 2004*, 65-72, Espoo, Finland, November 2004.

9 Samfat, D, Molva, R, Asokan, N. Untracebility in Mobile Networks. *The First International Conference on Mobile Computing and Networking (ACM MOBICOM 95)*, Berkely, California, USA, November 1995.

10 Go, J, Jim, K. Wireless Authentication Protocol Preserving User Anonymity. *The 2001 Symposium on Cryptography and Information Security (SCIS 2001)*, Oiso, Japan, January 2001.

11 3GPP. *3G Security; Wireless Local Area Network (WLAN) Interworking Security*. Sophia Antipolis, France, 2007. (3G TS 33.234)

12 Kent, S, Seo, K. *Security Architecture for the Internet Protocol*. IETF, December 2005. (RFC 4301)

13 Dolev, D, Yao, A. On the Security of Public-Key Protocols. *IEEE Transactions on Information Theory*, 2 (29), 1983.

14 Gong, L. Lower Bounds on Messages and Rounds for Network Authentication Protocols. In: *Proceedings of 1st Conf. on Computer and Comm. Security '93*, 1993.

15 Menezes, A J, van Oorschot, P C, Vanstone, S A. *Handbook of Applied Cryptography* (5th printing). CRC Press, June 2001. (ISBN 0-8493-8523-7)

16 Lauter, K. The Advantages of Elliptic Curve Cryptography for Wireless Security. *IEEE Wireless Comm. Magazine*, 11 (1), 62-67, 2004.

17 Burrows, M, Abadi, M, Needham, R. *A Logic of Authentication*. California, USA, 1990. (DEC SRC Research Report 39)

17 Handschuh, H, Preneel, B. *Minding Your MAC Algorithms*. Input paper to IST-2002-507932 (ECRYPT), 2004.

18 ECRYPT. *ECRYPT Yearly Report on Algorithms and Keysizes (2005)*. IST ECRYPT NoE (IST-2002-507932), Deliverable D.SPA.16, January 2006.

20 Hofheinz, D, Müller-Quade, J, Steinwandt, R. Initiator-Resilient Universally Composable Key Exchange. *ESORICS 2003*, Gjøvik, Norway, LNCS 2808, 61-84. Springer-Verlag, 2003.

21 3GPP. *Mobile radio interface layer 3 specification, Core network protocols; Stage 2*. Sophia Antipolis, France, 2005. (3G TS 23.108)

22 3GPP. *Mobile Application Part (MAP) specification*. Sophia Antipolis, France, 2007. (3G TS 29.902)

23 Kesdogan, D, Federrath, H, Jerichow, A, Pfitzmann, A. Location Management Strategies Increasing Privacy in Mobile Communication. *IFIP SEC 1996*, Samos, Greece, 1996.

24 Federrath, H, Jerichow, A, Pfitzmann, A. MIXes in Mobile Communication Systems: Location Management with Privacy. In: *Proc. First Intern. Workshop on Information Hiding*, Cambridge, UK, LNCS 1174, Springer, May/June 1996.

25 Asokan, N. *Anonymity in a Mobile Computing Environment*. Workshop on Mobile Computing Systems and Applications, December 1994.

26 Ateniese, G, Herzberg, A, Krawczyk, H, Tsudik, G. Untraceable Mobility or How to Travel Incognito. *Computer Networks journal*, 31 (8), 785-899, 1999.

For a presentation of the author, please turn to page 3.

Location Hidden Services and Valet Nodes

LASSE ØVERLIER, PAUL SYVERSON

Lasse Øverlier is a research scientist at the Norwegian Defence Research Establishment (FFI) and a PhD student at Gjøvik University College, Norway

Paul Syverson is a Mathematician at the Center for High Assurance Computer Systems (CHACS) of the Naval Research Laboratory (NRL), Washington DC, USA

Location hidden services are constructed to hide the physical location of a server carrying a publicly accessible service. This is of increasing importance nowadays with the ongoing discussions of privacy vs. surveillance. We will here briefly describe how these location hidden services work, how they are vulnerable and how we have reduced these vulnerabilities. We introduce so-called *valet nodes* to assist in protecting the contact points of the service. Our new protocol extension also enables the hidden service to exist without its presence being known to unauthorized users. In addition we will present how to add quality of service as a service option.

1 Introduction

Location hidden services are built to achieve high availability, making the service harder to attack both physically and logically. We have used Tor as the anonymizing network exemplifying the issues found in the current design of hidden services. These services are not specific to the Tor [9] network, but Tor have since 2004 been an anonymizing network implementing these hidden services.

Lately hidden services have increasingly been given more attention as a means to resisting denial-of-service (DoS) attacks, but most often the focus is on the challenge of locating (finding the IP address) of a specific service. The media have started using location hidden services as an example of how to circumvent censorship, for example by using tor-casting[1)], and to preserve freedom of speech. These may be issues both where dissidents suffer under a high degree of censorship, and in other arenas where for instance companies fire bloggers for having the 'wrong' opinion.

We have previously addressed [17] the vulnerabilities of locating these hidden services and methods to improve the service by being less vulnerable to location attacks like statistical attacks [7, 15]. But in [18] we address another important issue, resistance against DoS attacks and an improved flexibility in the service. The current hidden service design uses publicly known introduction points for being accessed, and this opens up for attacks simply on the basis of knowing these nodes. We will here give an overview of the improvements that can be made by using *valet nodes* to hide these introduction points, and the methods used to add quality of service (QoS) to a hidden service connection. A more detailed look at the protocol can be found in [18].

In Section 2 we present previous work on hidden services with a brief look into how Tor's hidden services work. In Section 3 we describe the introduction of the valet node into the protocol. In Section 4 we discuss the security of the design, and in Section 5 we present our conclusions.

2 Previous Work

The Eternity service [1] by Ross Anderson was the first system to hide the location of a service. It stored files at multiple locations for a certain period of time. GNUnet [2] communication builds upon mix-net [5] technology for sending messages to other nodes and is vulnerable in the way that it trusts the availability of the underlying network. Other systems with focus on publishing information are Freenet [6] – using a peer-to-peer network, Publius [14] – focus on persistence of stored files, Tangler [22] – making newly published files dependent on previous ones, and Free Haven [8] – using a reputation system between nodes for trust.

Location hidden services in the current design are made possible in low-latency anonymizing networks like Tor [9], a newer version of the earlier onion routing [11] communication protocol. Other low-latency systems are the Freedom Network [4] and JAP [3]. JAP is based on mix cascades with common entry and exit points, and will not be usable for the current hidden services design. Tor uses public-key cryptography to distribute session keys establishing circuits with perfect forward secrecy along a route through the network. Each session key is shared between the circuit initiator and each of the nodes along the path. When data is transported through the circuit a layer of encryption is added/removed for each hop making the data appear different at every node.

1) *Web/podcasting over Tor.*

Single node proxies like the Anonymizer [12] hide their client's connections through high amounts of traffic mixed in a single relay, and do not support hidden services at all. Single node proxies also suffer from the vulnerability of being a single point of failure, a single point of compromise, and a single point of attack.

Hidden services as described in [9] is not a publishing service in itself, but enables a service provider to run any TCP-based service, for example a web server or login server, through the anonymizing network so the location of the server is hidden from both the network and from the users of the service. This section describes the functionality of hidden services in Tor, and is taken from the articles [17] and [18].

2.1 Location-hidden Services in Tor

In the current implementation of Tor, a connection to a hidden service involves five important nodes in addition to the nodes used for basic anonymous communication over Tor:

- HS, the hidden server offering some kind of (hidden) service to the users of the Tor network, for example web pages, mail accounts, login service, etc.;
- C, the client connecting to the hidden server;
- DS, a directory server containing information about the Tor network nodes and used as the point of contact for information on where to contact hidden services;
- RP, the rendezvous point is the only node in the data tunnel that is known to both sides;
- IPo, the introduction point where the hidden server is listening for connections to the hidden service.

A normal setup of communication between a client and a hidden service is done as shown in Figure 1. All the displayed connections are anonymized; that is, they are routed through several anonymizing nodes on their path towards the other end. Every arrow and connection in the figure represents an anonymous channel consisting of at least two intermediate nodes. (Hereafter, we use 'node' to refer exclusively to nodes of the underlying anonymization network, sometimes also called 'server nodes'. Although we are considering the Tor network specifically, the setup would apply as well if some other anonymizing network were used to underlie the hidden service protocol. Unlike the other principals above, C and HS may be anonymization nodes or they may be merely clients external to the anonymization network.)

First the hidden server connects (1) to a node in the Tor network and asks if it is OK for the node to act as an introduction point for his service. If the node accepts, we keep the circuit open and continue; otherwise HS tries another node until successful. These connections are kept open forever; that is, until one of the nodes restarts or decides to pull it down.[2)] Next, the hidden server contacts (2) the directory server and asks it to publish the contact information of its hidden service. The hidden service is now ready to receive connection requests from clients.

In order to communicate with the service the client obtains a special URL, a *.onion* address, that it can understand and that has been posted to a public site or otherwise given to the client out-of-band. The client then connects (3) to DS and uses the *.onion* address to ask for the contact information of the identified service and retrieves it if it exists (including the addresses of introduction points). There can be multiple introduction points per service. The client then selects a node in the network to act as a rendezvous point, connects (4) to it and asks it to listen for connections from a hidden service on C's behalf. The client repeats this until a rendezvous point has accepted, and then contacts (5) the introduction point and asks it to forward the information about the selected RP.[3)] The introduction point forwards (6)

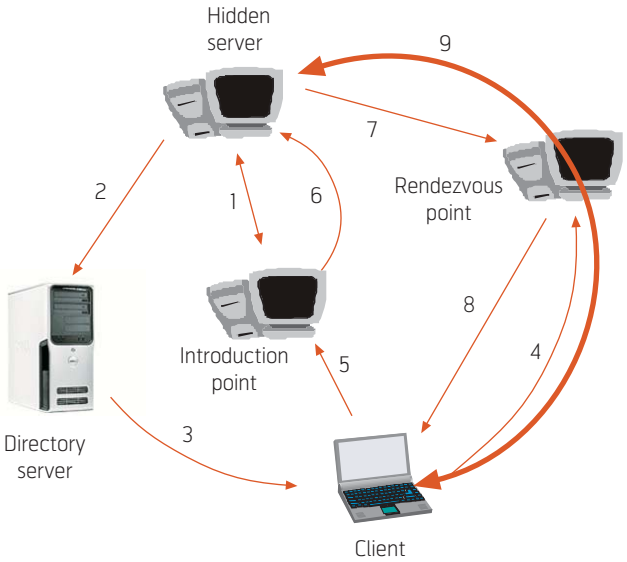


Figure 1 Normal use of hidden services and rendezvous points

2) *In Tor any node in a circuit can initiate a circuit tear down.*

3) *Optionally, this could also include authentication information for the service to determine from whom to accept connections.*

this message to the hidden server, which determines whether to connect to the rendezvous point or not. If OK, the hidden server connects (7) to RP and asks to be connected to the waiting rendezvous circuit, and RP then forwards (8) this connection request to the client.

Now RP can start passing data between the two connections and the result is an anonymous data tunnel (9) from C to HS through RP.

2.2 Threats to Hidden Services

Most papers addressing the security of anonymizing networks have focused on the threats of locating the services and their clients using attacks like intersection attacks [20, 24] and traffic analysis [19, 13, 25]. All low-latency networks are vulnerable to correlation of traffic between two suspected communicators, which means that the network is also vulnerable to an adversary controlling large portions of the network. In [21] Serjantov and Sewell show how a smaller adversary can match timing to and from nodes in the network. Murdoch and Danezis [16] attacked location through setting up communication with all nodes of the network and detected which nodes were affected when they pushed large amounts of traffic through the network. In [17], we demonstrated how to implement intersection attacks in a live anonymizing network and demonstrated that it was possible to locate a hidden service within minutes using only a single malicious node inside the network. We also described how to use other attacks to enhance the previously known intersection attack by using two nodes or (ab)using information publicly available in the network.

In this paper we focus on *availability and flexibility*, not *location*, as this is likely to be a next step for an attacker. If the attacker cannot locate the service, he might still be able to stop the service from being used.

One of the most vulnerable parts of the hidden service design is the publication and aggregation of information about *all* the hidden services available in the directory service. This information contains the name of the hidden service, the .onion address, and all contact information – including the list of introduction points, enabling an attacker to launch a DoS attack on these specific servers and effectively stop any access to the service itself. Censoring authorities may also use this list to stop the nodes listed as introduction points for one or more specific services. A node may also itself refuse to be used as an introduction point for any service it disapproves of and can therefore assist in reducing the service's availability either by the node's own decision or by being forced to do so.

And since all hidden services are listed at the directory service, they are not hidden from the directory server nodes. Even if these nodes are trusted not to abuse this information, they are still a single point of failure and attack. Just knowing about a service's existence is a potential for abuse.

3 Valet Service

A valet service is a helper service using *valet nodes* implementing so-called reply onions [11] to hide the introduction points from the users of the service. By introducing these valet nodes we also make it simple to hide the service and its very existence from the directory servers, and we include QoS that could be tailored for each individual user of the service.

The *valet nodes* are random nodes of the network, selected to be able to contact one or more of the introduction points of the hidden service. These nodes will of course have selection criteria like uptime, stability, etc. In order to contact an introduction point the valet node needs contact information that we have put inside a *valet token*. This valet token is encrypted with the valet node's public key, so it may be read by that node only. The valet token and the identity of the valet node is put inside a *contact information ticket* that is given to the client. This contact information ticket will enable the client to talk to the introduction point without knowing which node this is, and will not contain information revealing to the valet node which service the client is accessing.

In short, the valet node enables the client to contact the introduction point without knowing the the introduction point's identity, and without revealing to the introduction point who the client is nor what service the client is accessing.

3.1 Using Valet Nodes

When using valet nodes to contact a hidden service we will have additional levels of security to protect the introduction points. The full details of the valet nodes protocol can be found in [18] and include message descriptions, timestamps, use of multiple valet nodes, key exchange information and optional authorization information.

Figure 2 shows how the use of valet nodes protects access to the introduction point, while keeping the protocol close to the original hidden service design.

As before, the hidden service first contacts (1) an introduction point and asks this to listen for specific connections. In addition the hidden service gives the introduction point a private service key to use to identify itself when this service is being accessed.

The public part[4] of this key is put inside the contact information ticket and is used by the client to verify the introduction point's authenticity. In this way the client will be unable to identify the introduction point as it is now *not* using its own private/public key pair.

Instead of publishing the introduction points directly, the hidden service now selects a valet node with permission to contact this introduction point. The identity of the introduction point is included inside the valet token together with an identifier that is encrypted with the public key of the introduction point. This identifier will allow the introduction point to associate the client connections with the correct connection to a hidden server and to confirm that the correct valet point is forwarding the connection request. These parameters and a timestamp are encrypted using the valet node's private key to form the valet token. The identifier and the identity of the valet node are given to the introduction point during its setup ensuring[5] that both factors must be in place to allow a connection. The identity of the valet node and the valet token is then put inside a contact information ticket that is given to the client. This could be done offline or through the directory service (cf. Section 3.2).

When the client then wants to connect to the hidden service, it must have access to a contact information ticket. As before the client starts by contacting (2) a rendezvous point and asks this to await a connection request. The client unpacks the contact information ticket, identifies the valet node and makes a tunnel (3) to the valet node, sending over the valet token and asking the valet node to extend the tunnel (4) to the introduction point. By using the public part of the service key pair located inside the contact information ticket, the client authenticates (5) the introduction point without knowing its real identity. Now, the client is able to forward encrypted information (6) to the hidden service as before, including which rendezvous point to use and optional information authorizing access to the hidden service.

The hidden service then determines whether to contact (7) the rendezvous point and complete the tunnel setup (8,9) or to drop the request.

3.2 Distributing tickets and QoS

One important aspect of using contact information tickets is the similarity to the use of HTTP-cookies [10]. If the client is given a contact information ticket

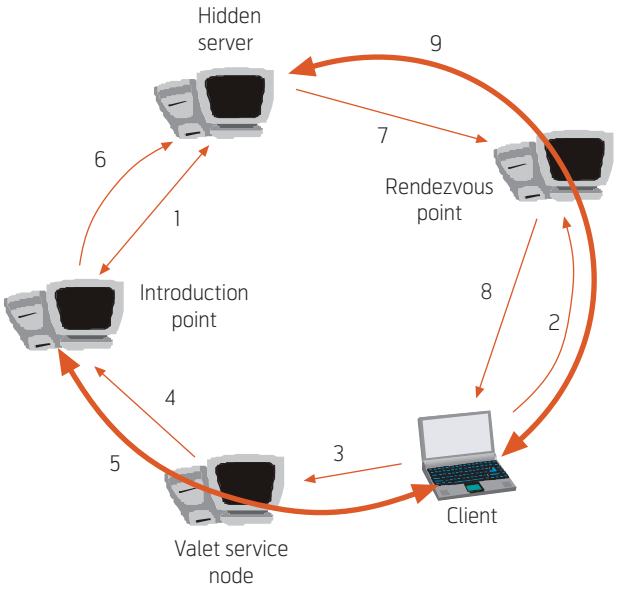


Figure 2 Use of Valet Service

directly and this ticket is, for example, updated during the communication, the hidden service is able to keep track of users and build a table of trust for a single user, or a group of users. If the client is authenticated through its use of the anonymized service this is similar to personalization of any type of service. But we are now able to build quality of service into the access of all users, both authenticated and anonymous. The hidden service could, for example, give more valet nodes and introduction points to the authorized users, decrease bandwidth for anonymous users, or even build reputation on an anonymous user giving her better (or worse) quality of service based on her previous behavior.

3.3 Hiding the Hidden Service

In the current design, all hidden services with all their respective introduction points are listed at the directory service. This means that all users of the network may find the introduction points to any service where they know the service's identifier, the *.onion* address. These identifiers are also known to the directory service which also is a potential threat as they will be able to contact and identify every hidden service running within the network; although by itself this does not allow either the directory service or arbitrary users to thereby locate the hidden server.

In order to contact the service the client must either have the service's public key, for example, identified in a contact information ticket, or the *.onion* address

---

4) *The terms* private *and* public *are used only to simplify the description as the introduction point holds its part of the key pair secret, while the other part is more accessible inside a ticket.*

5) *This verification can be done both by the introduction point and by the hidden service itself when the connection request arrives there, as the valet node has no reason to be kept anonymous from either the introduction point or the hidden service itself.*

to make a service lookup. In the first case the client does not need to make a service query as ticket control is maintained through the connections. If the client only has the identifying *.onion* address it must be able to make the lookup without revealing the service to others, including the directory service. We achieve this by using a simple scheme that rewrites the lookup identifier and encrypts the service information.

The client has (somehow) retrieved the *q.onion* identifier of the service to contact, where *q* is derived from the public key of the service, for example $q = hash(PK + value)$. We use this address, *q*, to create the service descriptor index, for instance $hash(q + \text{'1'})$, for retrieving the contact information ticket from the directory service. The downloaded value, *Q*, is the contact information ticket encrypted with a symmetric encryption scheme using a key derived from the public key, for example $hash(q + \text{'2'})$. So both the descriptor index and the descriptor content are hidden from the directory service, but available to any client that knows the *q.onion* address of the service. And finally, the verification of *q* may be completed using the public key and the signature of the contact information ticket as described above. There is no way to derive $hash(q + \text{'2'})$ from $hash(q + \text{'1'})$ without having *q*. Thus, the directory service cannot simply use the directory information it holds to contact the hidden service.

In order to create full flexibility for individual users, or groups of users, we may add a cookie inside the contact information ticket to be used during lookup and decryption, for instance use $hash(q + \text{'1'} + cookie)$ for lookup, and $hash(q + \text{'2'} + cookie)$ for decryption. Now we are also making it impossible for the directory service to count how many services it has listed because a service may have multiple entries. In addition, the cookies can be based on the client's authentication data, enabling only that specific client to download and decrypt the associated contact information ticket.

The scheme could also be expanded by using a date/time value inside the hash calculation to include a time period, for example current date/week, so a listing can exist anonymously without revealing when the service started to exist. Combining this with a time stamp could have the directory service store the entry for a longer (or shorter) period of time than default. And of course, for authenticated users we only need to give the client several contact information tickets with varying lifetimes. Typically any client should always have a longterm ticket and one or more short-term tickets.

3.4 Updating Contact Information Ticket

In order to verify an update of information inside the directory service during the entry's lifetime, we propose a simple reverse hash chain scheme where the initial contact to the directory service is followed by an iterated hash value, $v_n = hash^n(v)$, known only to the hidden service itself. For each update of this index (for example of $hash(q + \text{'1'} + date)$) the new encrypted ticket is accompanied by the value, v_{k-1}, enabling the directory service to verify the update using $v_k = hash(v_{k-1})$.

To adapt to the current and future improvements in hash collision techniques it is probably wise to increase the number of bits used in the *.onion* address from today's 80-bit (16 * base32(5-bit) characters) address to, for example, 256-bit (using 44 * base64[6](6-bit) characters, including an eight-bit version and extension value as the first byte of the address). An evaluation of hash algorithms will not be discussed here.

We now have a framework enabling the hidden service to be completely private and unavailable even to the directory service, able to track individual users, or groups of users, and give them different quality of service based on their local preferences.

4 Valet Node Security

Here we give a brief introduction to the different threats towards security in the old hidden service protocol, and how we have addressed these in our new design [18]. The most important threats are availability, man-in-the-middle attacks, denial-of-service attacks, and colluding nodes. We also look at the new threats introduced by the contact information tickets, and optional quality of service added to the connections.

4.1 Availability

By using valet nodes we have removed knowledge of the identity of the introduction points from the other nodes of the system: there is no way for a node knowing the *q.onion* address to locate the introduction points of this service. The node may get information about a valet node or two, but there may be many valet nodes, and many introduction points, so the problem of how to locate all introduction points is probably the most critical availability improvement in the new design. If the number of valet nodes is huge, we make it easier for an adversary controlling a part

6) *Use a modified base64, for example '/' → '-', '+' → '_'.*

of the network to identify the introduction points, as it is more likely to have control over one of the valet nodes in every introduction point's associated 'group' of valet nodes.

Using n as the number of nodes in the network, c as the number of compromised nodes, i as the number of introduction points, and v as the number of valet nodes per introduction point, we can get an expression for the probability of revealing all the introduction points to the adversary. The probability P_s for a specific combination of c compromised nodes in $i + 1$ groups is given by Equation 1.

$$P_s(x_j \text{ in Valet group } j) = \frac{\binom{G_0}{x_0}\binom{G_1}{x_1}\cdots\binom{G_i}{x_i}}{\binom{n}{c}} \quad (1)$$

Here 0 is the index for being outside all the valet node groups; that is, $G_0 = n - i \cdot v$, $x_0 = c - x_1 - x_2 - ... - x_i$, and all other G_j are given to be v. The number of compromised nodes c must be larger than or equal to the number of introduction points i, otherwise the probability will be zero.

$$P(n, c, v, i) = \sum_{x_1, x_2 \cdots, x_i} \frac{\binom{n-i\cdot v}{c-\sum_1^i x_j}\binom{v}{x_1}\cdots\binom{v}{x_i}}{\binom{n}{c}} \quad (2)$$

The probability of the adversary having concurrent presence in all groups is given by Equation 2 where $c \geq i$ and we sum over the values: $x_1 = 1, ..., \min(v, c - i + 1)$; $x_2 = 1, ..., \min(v, c - i - x_1 + 2)$; $x_3 = 1, ..., \min(v, c - i - x_1 - x_2 + 3)$; up to $x_i = 1, ..., \min(v, c - \sum_{j=1}^{i-1} x_j)$, where the upper limit of the x-values is v as indicated.

Figure 3 shows the probability that an adversary, controlling c nodes of the network, is able to locate all the i introduction points when using valet nodes, each using v valet nodes. Recalling from Section 2.2 that in the current design all nodes are known to everyone, this is a significant improvement to availability.

The more valet nodes added per introduction point, the higher the probability of locating all (via presence in all groups), and if we add more introduction points keeping the number of valet nodes constant, the probability decreases. We observe that the number of valet nodes is a more significant factor than the number of introduction points. For example, the strongest protection occurs in the case of using only a single valet node per introduction point, but this will, as previously mentioned, affect the service's availability. We also observe that when using nine introduction points and only one valet node per introduction point, the adversary will have to control around 400 nodes in order to have the same 10 % probability of locating them all.

In Figure 4 we compare the relative distributions of a network of 100 and 1000 nodes and observe only tiny variations in the probability distribution caused by the changing relative sizes of i and v compared to c and n.

Based on this we estimate that good protection of the service should consist of at least three introduction points combined with at least two valet nodes per introduction point, and should be combined with the possibility of differentiated quality of service as described in Section 4.4.

4.2 Man-in-the-Middle Attack

We have already addressed how the valet node is unable to perform a man-in-the-middle attack as the client authenticates and performs a key exchange directly with the introduction point based on a public key given in the contact information ticket. In addition the current design is also resistant towards a man-in-the-middle attack because of the key exchange between the client and the hidden service itself, a key exchange that takes place also in our valet node's extension.

4.3 Denial-of-Service Attacks

Contacting a valet node multiple times with the same token in order to deplete resources by forcing multiple decryption of tickets can be countered by having the token cached and thus only decrypted once. The valet node can also act only once per circuit on request to extend to an introduction point, forcing the attacker to build a new tunnel for each attempt. And the introduction point could also close every attempt and force the client to build a completely new tunnel through the valet node for each connection request.

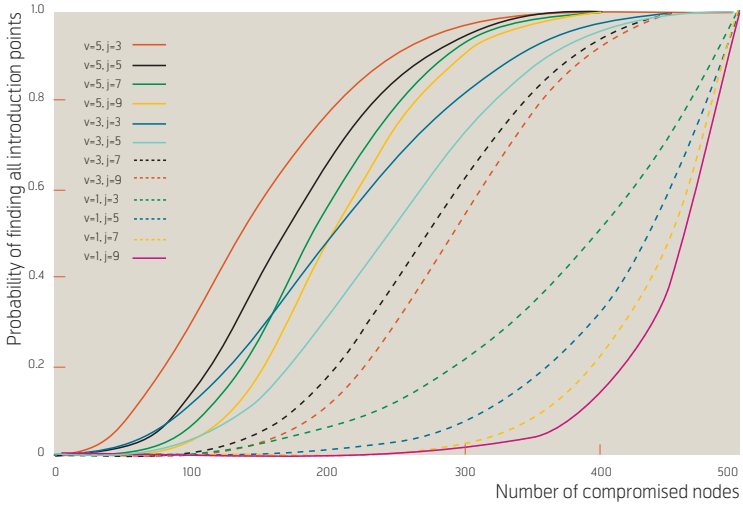


Figure 3 Probability of finding all i *Introduction Points, each using* v *Valet nodes, in a network of* n = *500 nodes*

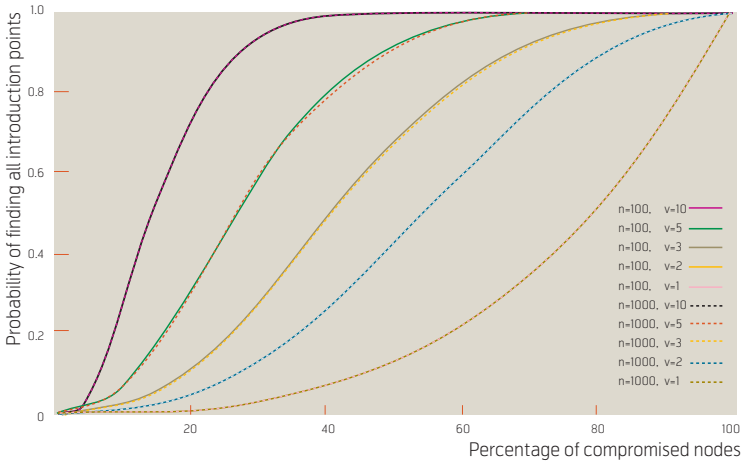


Figure 4 Comparing n = *1000 to an up-scaled* n = *100 for a service using three Introduction Points with varying number of Valet nodes*

In addition the underlying TLS[7)] connection from the valet node to the introduction point could stay open for a longer period of time than regular inter-server connections, making this attack quite useless. This counters hostile denial-of-service attacks but increases overhead for both the client and the network in the event of benign errors.

If the valet node is unavailable, the client must choose another valet node from the ticket. If they are all unavailable, this will affect all clients using the same ticket, and the client must find a way to retrieve another ticket. This vulnerability may be reduced by letting the clients have one or more long term tickets, or anonymous access tickets. The same problem occurs if the introduction points are unavailable, but this problem has not changed significantly from the current design.

4.4 Quality of Service

If a user wants to stay anonymous and untraceable, he must start with a public ticket every time (the paranoid variant), or trust the service to supply semi-public tickets for every connection, which, for example, the user can check by connecting multiple times using the public ticket(s). As these are open public services connected to by anonymous clients, this is an easy and simple verification.

When it comes to authorized users, a service may access the rendezvous point through different nodes giving a specific ('deserved') bandwidth to the user. This might reduce the set of nodes the service is selecting from when setting up the tunnel.

4.5 Colluding Connection Nodes

Alice may collude with the valet node and/or the introduction point. If she controls both, we have the same scenario as in the current design when Alice owns an introduction point. If Alice controls only the valet node, she will be able to find one of the introduction points, but the others will maintain availability for the service. If Alice controls an introduction point, she will not be able to verify this without either controlling the valet node or having access to the service's contact information ticket. Even when such a match is done this will give Alice no advantage compared to the existing system.

4.6 Contact Information Ticket Security

By using tickets we also introduce new challenges into the protocol.

Lost tickets, ticket expiration and valet node unavailability could be addressed by having longterm tickets. After authentication of the user (or user group) is completed, the quality of service may be upgraded (or downgraded) to the correct level for the user.

When a service has a secret hidden key (*.onion* address) to hide its existence, there is the problem of what happens when the address becomes public (or lost). Of course, if the service uses an authentication protocol for access, the problem is only confirmation of the service's existence and the possibility to target the service with attacks like the ones described above. This problem may be solved if we add the option of redistributing new public keys with the ticket update for authenticated users, or the service redistributes new tickets and service identifiers outside the directory service.

We also introduce the problem of 'free storage' into the directory service. For known services this is easy to counter by signing, but for hidden services with encrypted tickets this is an issue. We proposed a reverse hash chain scheme (Section 3.2) to counter false updates, but this does not stop the storage problem. We therefore propose that the insertion of data into the directory service is 'paid for', for instance by solving a computational puzzle of some kind.

5 Conclusion

We have presented an extension of the current hidden service design that improves availability and resistance to denial-of-service attacks through the introduction of valet nodes – nodes that hide the service's introduction points. The new design also facilitates

7) *TLS (Transport Layer Security) is the technology used for encrypting secure web connections.*

the use of 'completely hidden services': only clients that know a hidden service's *.onion* address or its public key will be able to connect to it or even verify the service's existence. The new protocol also allows differentiation of the quality of service given to clients, regardless of whether they are anonymous or authenticated.

References

1 Anderson, R J. The eternity service. In: *Proceedings of Pragocrypt '96*, 1996.

2 Bennett, K, Grothoff, C. GAP – practical anonymous networking. In: Dingledine, R (ed). *Proceedings of Privacy Enhancing Technologies workshop (PET 2003)*. Springer, March 2003. (LNCS 2760)

3 Berthold, O, Federrath, H, Köpsell, S. Web MIXes: A system for anonymous and unobservable Internet access. In: Federrath, H (ed). *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*. Springer, July 2000, 115–129. (LNCS 2009)

4 Boucher, P, Shostack, A, Goldberg, I. *Freedom systems 2.0 architecture*. White paper, Zero Knowledge Systems, Inc., December 2000.

5 Chaum, D. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 4 (2), 1981.

6 Clarke, I, Sandberg, O, Wiley, B, Hong, T W. Freenet: A distributed anonymous information storage and retrieval system. In: *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, July 2000, 46–66.

7 Danezis, G. Statistical disclosure attacks: Traffic confirmation in open environments. In: Gritzalis, Vimercati, Samarati, and Katsikas (eds). *Proceedings of Security and Privacy in the Age of Uncertainty (SEC2003)*, Athens, May 2003, 421–426. Kluwer. (IFIP TC11)

8 Dingledine, R, Freedman, M J, Molnar, D. The Free Haven Project: Distributed anonymous storage service. In: Federrath, H (ed). *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*. Springer-Verlag, July 2000. (LNCS 2009)

9 Dingledine, R, Mathewson, N, Syverson, P. Tor: The second-generation onion router. In: *Proceedings of the 13th USENIX Security Symposium*, August 2004.

10 Fielding, R et al. *Hypertext transfer protocol – http/1.1*. June 1999. (IETF RFC 2616)

11 Goldschlag, D M, Reed, M G, Syverson, P F. Hiding Routing Information. In: Anderson, R (ed). *Proceedings of Information Hiding: First International Workshop*, May 1996, 137–150. Springer. (LNCS 1174)

12 *Anonymizer Inc*. http://www.anonymizer.com/

13 Levine, B N, Reiter, M K, Wang, C, Wright, M K. Timing attacks in low-latency mix-based systems. In: Juels, A (ed). *Proceedings of Financial Cryptography (FC '04)*. Springer, February 2004. (LNCS 3110)

14 Rubin, A D, Waldman, M, Cranor, L F. Publius: A robust, tamper-evident, censorship-resistant, web publishing system. In: *Proceedings of the 9th USENIX Security Symposium*, August 2000, 59–72.

15 Mathewson, N, Dingledine, R. Practical traffic analysis: Extending and resisting statistical disclosure. In: *Proceedings of Privacy Enhancing Technologies workshop (PET 2004)*, May 2004. (LNCS)

16 Murdoch, S J, Danezis, G. Low-cost traffic analysis of Tor. In: *Proceedings of the 2005 IEEE Symposium on Security and Privacy*, IEEE CS, May 2005.

17 Øverlier, L, Syverson, P. Locating hidden servers. In: *Proceedings of the 2006 IEEE Symposium on Security and Privacy*, IEEE CS, May 2006.

18 Øverlier, L, Syverson, P. Valet services: Improving hidden servers with a personal touch. In: *Proceedings of the Sixth Workshop on Privacy Enhancing Technologies (PET 2006)*, Cambridge, UK, June 2006. Springer.

19 Raymond, J-F. Traffic Analysis: Protocols, Attacks, Design Issues, and Open Problems. In: Federrath, H (ed). *Proceedings of Designing Privacy Enhancing Technologies: Workshop on Design Issues in Anonymity and Unobservability*, 10–29. Springer-Verlag, July 2000. (LNCS 2009)

20 Reiter, M, Rubin, A. Crowds: Anonymity for web transactions. *ACM Transactions on Information and System Security*, 1 (1), 1998.

21 Serjantov, A, Sewell, P. Passive attack analysis for connection-based anonymity systems. In: *Computer Security – ESORICS 2003*, October 2003.

22 Waldman, M, Mazières, D. Tangler: a censorship-resistant publishing system based on document entanglements. In: *Proceedings of the 8th ACM Conference on Computer and Communications Security (CCS 2001)*, November 2001, 126–135.

23 Wright, M, Adler, M, Levine, B N, Shields, C. An analysis of the degradation of anonymous protocols. In: *Proceedings of the Network and Distributed Security Symposium – NDSS '02*, IEEE, February 2002.

24 Wright, M K, Adler, M, Levine, B N, Shields, C. The predecessor attack: An analysis of a threat to anonymous communications systems. *ACM Trans. Inf. Syst. Secur.*, 7 (4), 489-522, 2004. A preliminary version of this paper appeared in [23].

25 Zhu, Y, Fu, X, Graham, B, Bettati, R, Zhao, W. On flow correlation attacks and countermeasures in mix networks. In: *Proceedings of Privacy Enhancing Technologies workshop (PET 2004)*, LNCS, May 2004.

Lasse Øverlier received his MSc degree from the Norwegian Institute of Technology (NTH), University of Trondheim, in 1993. From 1994 to 1995 he was a scientist at the Norwegian Computing Centre (NR) and from 1995 to 2001 he worked as technical manager in EUnet Media AS. He has from 2002 worked as a research scientist at the Norwegian Defence Research Establishment (FFI), and has the last three years been able to work towards a PhD at Gjøvik University College within the field of anonymity networks. He has since 2002 also been an employee of Gjøvik University College and lectured various network security classes. His main areas of research are privacy, network anonymity, hidden services, RFID, network- and software security.

email: lasse@hig.no

Paul Syverson is inventor of Onion Routing, for which he received the Edison Invention Award, and designer of all three generations of Onion Routing systems, including the latest system, Tor. Dr. Syverson has been designing and analyzing security and privacy systems at the Naval Research Laboratory for eighteen years. He has been chair of eight conferences and workshops ranging from the European Symposium on Research in Computer Security to the Privacy Enhancing Technologies Workshop and the Financial Cryptography Conference. He is editor of several books on these topics, as well as author of many dozens of papers published in refereed conferences and journals. He is also the author of Logic, Convention, and Common Knowledge, a book that discusses philosophical foundations of logic, and employs game theory and distributed computing in doing so. He is former editor of IEEE Cipher.
He has been an invited visitor at the Newton Institute for Mathematical Sciences in Cambridge, England and was on the faculty of the first International School on Foundations of Security Analysis and Design in Bertinoro, Italy. Degrees: PhD and MA in philosophy (logic), MA in mathematics (all three from Indiana), AB in philosophy from Cornell.

More information available at: www.syverson.org

A Framework for Efficient Security and Privacy Solutions in Data Intensive Wireless Sensor Networks

VLADIMIR ZADOROZHNY, VLADIMIR A. OLESHCHUK, PRASHANT KRISHNAMURTHY

Vladimir Zadorozhny is Assistant Professor at University of Pittsburgh, USA

Vladimir A. Oleshchuk is Professor of Computer Science at Agder University College, Norway

Prashant Krishnamurthy is Associate Professor at University of Pittsburgh, USA

We consider a rigorous framework for handling security and privacy solutions in Data Intensive Sensor Networks (DISNs) based on reasoning about data integrity constraints using application-driven data dependencies. The proposed approach provides a systematic and efficient way for privacy preserving data management and secure data distribution that compliment existing methods and techniques implemented on lower network layers. In particular we demonstrate how our framework can be combined with lower level wireless sensor network features such as collision handling.[1)]

1 Introduction

Wireless Sensor Networks (WSNs) naturally apply to a broad range of applications that involve system monitoring and information tracking (for example airport security infrastructure, monitoring of children in metropolitan areas, product transition in warehouse networks, fine-grained weather/environmental measurements, medical purposes etc.). Meanwhile, there are many open research issues in applying existing WSNs when the applications have high bandwidth needs for data transmission and stringent delay constraints against the network communication. Such requirements are common for *Data Intensive Applications (DIAs)* that utilize wireless sensor networks for managing critical assets. For an example of such DIA consider the task of Structural Health Monitoring (SHM) concerned with monitoring the integrity of civil and military structures, or a wireless sensor network for monitoring vital signs parameters from patients in a metropolitan environment, where situations of data loss, corrupted or delayed data can have implications for life and death. Another example of a *DIA* is the near-continuous monitoring (every two seconds) of heat exchangers in a nuclear power plant. Because the location of the wireless sensor nodes can be inaccessible per nuclear plant regulations, changing batteries frequently is not an option.

Managing critical assets requires high security guarantees which are difficult to achieve in wireless sensor nets. For example, secure methods for data interrogation in wireless sensor networks are required to avoid open transmitting of the critical structural, or patient health information. Meanwhile, most security protocols and architectures have been primarily designed for wired networks and in comparison only a little work is available on addressing security efficiently in wireless sensor networks.

Security in sensor networks encompasses several different aspects as outlined below.

- *Privacy of data:* Only authorized parties should be able to recover and understand data delivered by sensors.

- *Privacy of location/activity of sensors:* Only authorized parties should be able to discover the location and activity of sensors.

- *Secure data aggregation:* When data from several sensors are delivered to a sink, false, spoofed, or malicious data must not affect the data quality.

- *Denial-of-service:* Malicious actions may hinder delivery of data in a timely manner to the sink or destination. Alternatively, the lifetimes of sensors may be reduced by battery exhaustion attacks.

Measures undertaken to secure sensor networks should include (a) prevention mechanisms, (b) detection mechanisms, (c) response, and (d) assessment. Traditional prevention mechanisms apply cryptography as a tool to ensure privacy of data and/or eliminate spoofed or malicious data. Cryptography however is computationally extensive and affects the battery life of sensors [Djen05] as well as creating latency in transmitting data. Previous work by the authors [Pras04] has shown that the size of transmitted packets, the encryption/authentication algorithms, key sizes can impact the latency and energy consumption in wireless networks. Moreover, asymmetric or public key schemes can consume as much or more battery power as the transmission of a packet requiring complex key management schemes that can increase congestion in sensor networks. Detection mechanisms employ monitoring of traffic, relaying of alarms, and cryptography, which again impacts the battery life of sensors. Response mechanisms in general are reactive and have strong dependence on the detection schemes. Finally, very little assessment of security is performed in sensor networks. In summary, there is a need for interaction between security

1) *This work is supported in part by the Norwegian Research Council under the BILAT project 174958/D15.*

and privacy policies, security protocols and low-level security primitives to achieve the highest efficiency for a desired degree of security.

In this paper we consider a general framework for handling security and privacy solutions in Data Intensive Sensor Networks (DISNs). Our framework is based on observing and reasoning about data integrity constraints using application-driven data dependencies. The proposed approach provides a systematic and efficient way for privacy preserving data management, secure data distribution, integration and searching that compliment existing methods and techniques implemented on lower network layers. The paper is organized as follows. In the next section we overview existing privacy and security solutions in wireless sensor networks. Section 3 elaborates on the proposed approach for security and privacy preservation based on semantic data dependencies. Section 4 describes how our data semantic approach can be combined with lower level wireless sensor network features such as collision handling. Section 5 concludes and outlines further research directions.

2 Overview of Existing Privacy and Security Solutions in Wireless Sensor Networks

Security in wireless sensor networks is unique in that the threats that need to be considered are quite different from those in other kinds of networks or distributed systems. Sensor networks comprise of perhaps thousands of low cost, battery-operated units with limited computational power and memory that need to communicate potentially with one another and some sink or base station using wireless links. Consequently, scalability (because of the sheer number of nodes), energy efficiency, necessity of lightweight cryptographic protocols, and methods to overcome wireless vulnerabilities are important. Moreover, it is not possible to assume that sensor units will not be compromised or tampered with for the same reasons. Hence, it is important to consider both outsider and insider threats in sensor networks. In the following sections, we describe security threats to wireless sensor networks and security measures proposed in the research literature. The interested reader is referred to [Djen05] for a survey of security issues in the superset of ad hoc networks, to [Zhe06] for security in low rate wireless networks, and to [Shi04] for a discussion of design issues for security in sensor networks.

2.1 Threats in Wireless Sensor Networks

There are numerous security threats in wireless sensor networks making it difficult to consider them all together. Instead it is easier to group the threats into categories. While there are overlaps between them, we can classify these threats into the following categories (see Figure 1).

Physical layer threats: At the physical layer of the communications protocol stack, common threats against sensor networks are disruptions to communications through jamming [Woo02] and node disabling. By jamming, an attacker may disrupt reliable communications by transmitting signals that interfere with the radio signals of sensor nodes. This may result in partitioning of the network, lower reliability of the sensed data because of the lack of availability of data from certain sensed regions, and ultimately result in the battery exhaustion of nodes repeatedly transmitting data till they are acknowledged or receiving bogus data. [Xu06] classifies jammers as those that may be outsiders employing a constant radio signal, deceptive jammers that inject regular packets into the network, random jammers that alternate between sleep and awake states, and reactive jammers that cleverly disrupt communications upon sensing channel activity. Experimental work by [Xu06] indicates that packet delivery ratios are adversely impacted by all of these types of jammers.

Eavesdropping threats: One of the most common threats in wireless sensor networks is information leakage [Shi05] where an adversary may obtain the

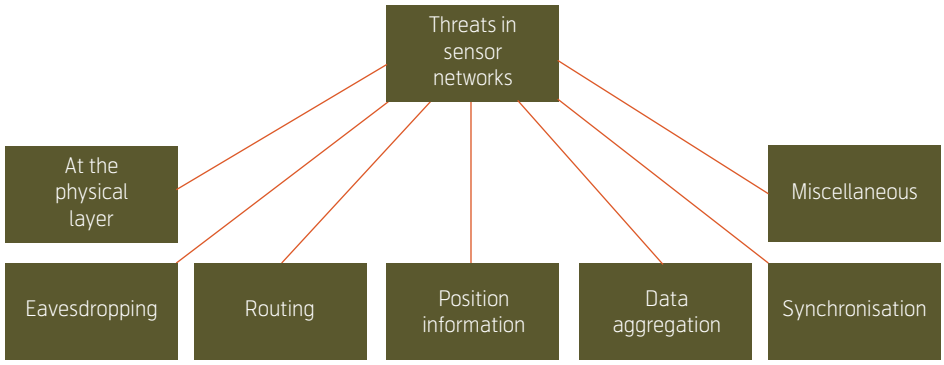


Figure 1 Security threats in sensor networks

sensed information by simply passively eavesdropping on the radio signals being transmitted by sensor nodes. Traditionally, cryptography has been used for protecting the confidentiality of information in networks. Cryptography can be computationally intensive, may require additional memory and needs scalable key management, which are all issues in wireless sensor networks. Eavesdropping is especially problematic even with encryption because of the potential for sensor nodes possessing keys to be compromised or captured by adversaries. In [Ana05], a model for computing the eavesdropping vulnerability is presented where the adversary is interested in predicting the behavior or aggregate output of the sensor network. In addition to information leakage, radio transmissions may reveal the location of sensors and the sink node and allow other kinds of analyses on the traffic patterns. An adversary may also be able to actively poll sensor nodes for information if there is no authentication of queries in the network.

Threats impacting routing: In networks, it is important for nodes to know where to send data packets so that they reach the destination in an efficient way. Such routing protocols in sensor networks are still evolving since attempts to directly use routing protocols designed for mobile ad hoc networks in sensor networks have faced challenges due to the scalability and energy requirements of sensor networks. Geographical and geometric routing that makes use of the knowledge of the Euclidean coordinates of sensors is proposed as an efficient means of routing data to the destination (see for example [Bru05]). However, it is likely that in general, routing in sensor networks faces the same threats as those in mobile ad hoc networks. Such threats include location disclosure, replay of old routing information, disruption by fabricating routing information, and route table poisoning [Arg05]. In addition, wormhole, blackhole, and Sybil attacks are possible.

In blackhole attacks, malicious nodes advertise themselves as closer to the destination thereby making themselves part of most routes. They can then disrupt network operation by dropping packets or get information by eavesdropping. In Sybil attacks [New04], a single malicious node claims to be more than one node. This way, it could claim a disproportionate amount of resources and also perform blackhole attacks. If there are collaborating nodes, they may create a wormhole (a tunnel) between them and create the impression of a false network topology.

Threats impacting position information: The position of a sensor node has importance in several applications. For example, temperature variations over a given area may have to be accurately characterized, in which case the position location of the sensor reporting the temperature reading needs to be known to a certain accuracy. Such position information may be used for routing or even in security measures (see for example [Dem06]). Further, the location of a sensor monitoring a critical quantity may itself need to be kept secure (location privacy). Malicious nodes can interfere with the reporting of position location information in many ways. They can fabricate the position information or interfere with the support infrastructure using which sensors can determine their positions. In the latter case, there are many different approaches for determining the position of sensor nodes such as using beacons from nodes at known positions, determining the number of hops a node is away from a reference node, and so on. Malicious nodes can interfere with such position determining activity.

Threats impacting data aggregation and in-network processing: An important part of many sensor networks is data aggregation and in-network processing. Because sensor nodes collect a huge amount of data and sometimes only aggregate information is necessary at the sink (for example, average value or the sum of the sensed quantity), intermediate nodes can process the received data (in-network processing) or fuse data and forward those values. This reduces the communication costs and delays in the network. However, such functionality makes it extremely easy for malicious nodes to introduce false values that corrupt the processed or fused values. If a malicious node is responsible for fusing or aggregating data, the problem could be worse. If a Sybil attack is launched, a node can claim multiple identities and further skew the aggregated data by creating multiple false reports.

Threats against time synchronization: Sensor networks often require nodes in the network to be time synchronized for many reasons such as data fusion, scheduled transmissions for saving power, tracking duplicate sensed data and so on [Siv04]. Time synchronization can be achieved using reference broadcasts or sender-receiver synchronization. It is possible to disrupt the sensor network operation by misleading different nodes about the time at which they have to perform operations like sensing or transmissions of packets.

Miscellaneous threats: If sensor nodes are compromised, they can disrupt a sensor network in many ways. For instance, a compromised node may not follow the medium access protocol and hog the medium. If a node assumes several identities (Sybil attack), it can access the medium more often than usual. It is common for each node to have fair and equal access to the medium. By pretending to be many nodes, such

an attacking node gets unfair access to the medium. These may both deny access to radio resources by legitimate sensor nodes. In many types of sensor networks, a sink node is used to collect data after a query to many sensor nodes in the networks and for other types of network maintenance. In some cases, mobile sink nodes are employed to poll sensors or collect data from a set of static sinks. Compromise of sink nodes can lead to damages or disruption of a sensor network.

2.2 Security Measures in Wireless Sensor Networks

Security measures undertaken to secure sensor networks should include (a) prevention mechanisms, (b) detection mechanisms, (c) response, and (d) assessment. Below we discuss some of the approaches reported in the literature. A summary of the discussion is illustrated in Figure 2.

Prevention mechanisms: Prevention mechanisms proactively try to eliminate potential threats to security. We can classify prevention schemes into those based on cryptographic protocols and those based on protocol measures intended for security.

Encryption is a well-known technique for concealing information from passive eavesdroppers thereby ensuring its privacy. If nodes in a sensor network are not compromised and they share secret keys appropriately, the sensed information can be kept secure. Even data aggregation and transmission to a central sink node can be kept secure by locally sharing keys within a subgroup of nodes (see for example [Wes06]). Sending spurious data, that is separable by the sink node but not by eavesdroppers, once again using encrypted IDs, is another way of concealing information [Ana05].

Much of the threats in sensor networks emanate from fabricated data or control signals. These include fabricated routing data, falsified position or reference position information or timing information intended for synchronization. Using cryptography for authenticating data is a well-known technique. Authentication makes use of either shared secret keys between parties for message authentication codes (MACs) or digital signatures of the sender. There are several cryptographic protocols making use of authentication to secure routing such as authenticated routing for ad hoc networks (ARAN), Ariadne, and secure ad hoc on demand distance vector routing (SAODV) [ARG05]. Authentication is also used for securing time synchronization messages – for example, the fault tolerant cluster-wise clock synchronization protocol in [Sun05] assumes that each pair of sensor nodes shares a unique secret key. Position information can also be kept secure by authenticating [Laz05] the reference information (beacons, node identities or positions, hop counts etc.).

Cryptography is computationally intensive and affects the battery life of sensors [Djen05] as well as creating latency in transmitting data. For example, [Pras04] has shown that the size of transmitted packets, the encryption/authentication algorithms, key sizes can impact the latency and energy consumption in wireless networks. The encryption overhead of many secret key schemes and hash functions in sensor nodes is reported in [Gan03]. Data and memory requirements of many block ciphers in a microcontroller have been reported in [Law06].

Key management [Heg06] is a problem when there are hundreds or thousands of nodes. If all nodes share the same key, the compromise of one node can render the cryptographic mechanism useless. If each node shares a pair of keys (unique pairwise keys), the number of keys to be installed, stored, and managed can become formidable. Asymmetric or public key schemes, typically used in wired networks to address such problems, can consume as much or more battery power as the transmission of a packet requiring complex key management schemes. Randomly distributing [Esc02, Cha03] a subset of keys in each sensor node has been proposed as one solution to this prob-

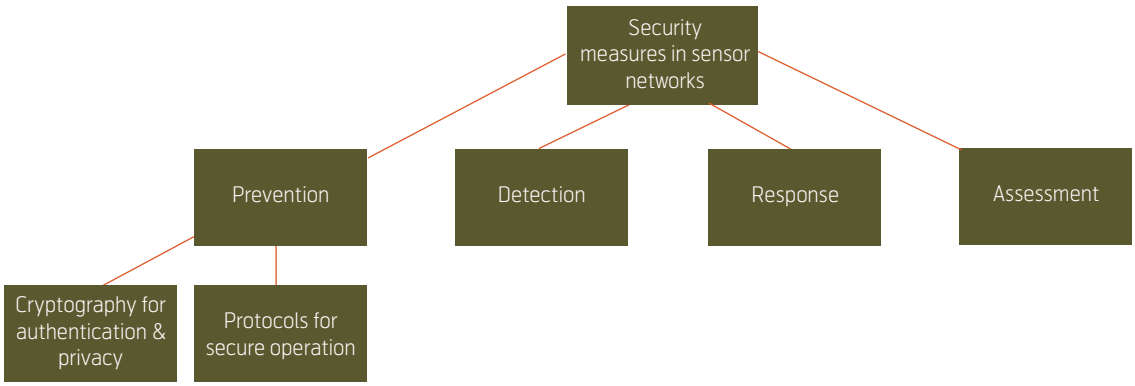


Figure 2 Security measures in sensor networks

lem. However, all nodes may not share keys with one another in this case. This can increase communications in the sensor network till nodes figure out with whom they share keys and how to reach a destination. There are also key management schemes that allow up to *k* nodes in the network to be compromised without rendering the whole network insecure [Du03].

Protocols can be used for prevention against security threats. Multipath routing has been suggested as a means of overcoming blackhole attacks. The information is routed along multiple paths to the sink so that even if malicious nodes are part of one route, transport through other routes is successful. A technique has been suggested for source location privacy that makes use of random walks to create phantom sources followed by delivery of messages to the sink [Zha06]. The idea here is that the data will appear to be delivered to the sink from geographical locations different from the point where the data has been sensed. Thereby, eavesdroppers cannot be sure of where the data was sensed. Careful consideration has to be given to protocols designed to prevent attacks as they may be exploited by adversaries to launch other types of attacks.

Detection mechanisms: Detection of active attacks is important in sensor networks especially because of the threat of compromised nodes. Detection of attacks requires active monitoring of activities of sensor nodes. Intrusion detection in sensor networks is different from that in wired networks as the corresponding threats and attacks are different. For instance, sensor nodes or other entities need to monitor packet forwarding to discover blackhole attacks. This is typically done by adding a *reputation* to each node based on its behavior (see for example [Mar00] and [Buc02]). If the reputation of a node drops below a threshold, it is marked as malicious.

Packet leashes (that estimate the distance a packet has traveled in a given amount of time) have been proposed as a mechanism to detect wormhole attacks [Arg05]. Using the location information and signal strength information to detect Sybil attacks has been proposed in [Dem06]. In [Xu06], the problem of distinguishing between adversarial jamming and environmental effects has been mentioned. The authors propose the use of basic statistics (signal strength, carrier sensing time and packet delivery ratio) and advanced detection strategies (combining these statistics) to detect jamming attacks.

A simulation of a decentralized intrusion detection system for sensor networks has been presented in [Sil05] where attacks such as jamming, delaying messages, selective forwarding, message alteration and wormhole attacks were considered and the detection effectiveness determined. The detection was based on previous data and application of rules. Using small amounts of data to detect attacks was found to yield false positives.

Response and assessment: It is common to have a standard set of responses when attacks are detected in wired networks depending on policies and procedures. This is usually followed by an assessment of reasons why attacks were successful. In sensor networks, we will consider as response to detected attacks, the mechanisms employed in the wake of successful detection of specific attacks. Very little work has been reported on assessment of the security of wireless sensor networks.

It is common to blacklist a node when its reputation has fallen below a threshold. For example, in [Mar00] a *Pathrater* is used to determine routes with the best reputation thereby eliminating routes that include nodes with bad reputation. If a node is marked as malicious, it is also possible to eliminate it from routing tables, disregard data sensed by such a node, or not use the information supplied by it for time synchronization or positioning purposes. The keys used by compromised nodes, when such nodes are detected, need to be revoked. This can be an expensive operation considering the number of nodes in the network and the total number of keys that need to be changed. An efficient key revocation protocol is presented in [Din06]. At the physical layer, many steps can be taken in response to jamming attacks. [Xu06] presents *evasive defensive strategies* where sensor nodes change channels. The entire network may change channels in response to a jamming attack. If nodes are mobile, they may leave the jammed area (spatial retreat). The nodes may also change their transmit power, reduce their code rate or modulation levels to improve the reliability of packets under jamming.

Code attestation has been suggested in [Shi04] as a method for assessing the validity of the code running on sensor nodes to ensure that nodes have not been compromised. If such assessment can be performed efficiently by a remote party using trusted hardware, the security of sensor networks may be further improved.

3 Security and Privacy Preservation Using Data Dependencies

In this section we consider an approach that utilizes application driven data integrity constraints for security and privacy preservation in wireless sensor networks. The integrity constraints are represented as

dependencies between sensed parameters extracted from application semantics. We start the explanation of our approach by considering a motivating example of a wireless sensor network for large-scale patient health monitoring. We will use this example to illustrate dependency-based reasoning for security preservation.

3.1 Motivating Example: Wireless Health Monitoring

As a motivating example we consider a wireless sensor network for monitoring vital signs parameters from patients in a metropolitan area. Such a network includes body sensors communicating with a receiver unit, for example a Hand Held Device (HHD) carried by a patient, which in turn can use another wireless hoop (for example GPRS telecommunication solution) in order to transfer data to a central base station. The HHDs can be bypassed when a patient is at home and the body sensors can transmit directly to a home-installed stationary wireless sensor HUB instead of using a wearable HHD. Depending on the actual situation the sensors should be provided with extra bandwidth. For example, when patient condition is normal body sensors can report their parameters infrequently. In case of an abnormal situation the sensors should automatically increase the data transmission rate and report health parameters more frequently. Under critical health conditions the sensor will need maximum bandwidth for near-to-continuous transmission of monitored parameters.

Privacy protection is critical in a wearable sensor network. Security precautions must be taken to ensure correct operation. For example, a body sensor will have to transmit an individual unique number (*Sensor_ID*) which has to be associated with the correct patient (*Patient_ID*). Both *Patient_ID* and *Sensor_ID* are required to assess patient condition with respect to the measured vital signs. Sensor readings associated with HHD or HUB can also reveal geographic location of the patient. Meanwhile, patient location should not be exposed under normal circumstances, while in critical situations both sensor location and the accurate values of the monitored parameters must be provided. For example, consider a patient monitored for his heart condition by a wireless ECG (Electro cardiography) sensor detecting abnormal heart beats and life threatening cardiac activities. If the sensor detects a sudden heart attack (ventricular fibrillation) it is time-critical to start a rescuing procedure. Thus, the sensor should be able to trigger an automatic escalation of the WSN, where this sensor is given a higher priority to ensure reliable data transmission. At the same time, this situation should open protected data to give the rescuing personnel more information about time, place and other relevant data transmitted from the area of accident.

To sum up, sensor networks transmit monitoring data via a wireless medium and are thus vulnerable with respect to privacy and security. Sensor measurements represent private information about monitored objects which implies that data transmissions and data flow within and out of the sensor network should be protected. At the same time, WSNs should provide efficient authorized access to measured data in order to perform monitoring and event detection. Since sensors have limited battery life-time, low data transmission rates and computational power, traditional privacy protection approaches based on strong encryption cannot be applied directly. On the other hand, in many cases application of cryptography is not necessary to archive monitoring goals. For example, the different degree of information leakage can be accepted as a result of a trade-off between requirements to privacy and performance. In order to reason about such trade-offs the WSN should utilize application semantics to decide what parameters should be transmitted and what levels of refinement in parameter values (e.g. actual data readings, or an aggregated estimate) should be provided. Automatic escalation algorithms should take this application semantics to choose an appropriate transmission mode, which can be implemented by either transmitting more parameters, or transmitting more refined values of some parameters.

In the next section we introduce a formalism that captures the relevant application semantics in order to provide above functionality. The proposed formalism supports efficient reasoning about the application semantics using parameter dependencies. We will demonstrate that our approach can be combined with a lower-level mechanism to ensure comprehensive privacy and security preservation in wireless sensor networks.

3.2 System Model

Generally speaking, sensor networks support distributed interaction with the physical environment through measuring and aggregation of data in order to create a dynamic global view. Various streams of measured data can be used to monitor and detect events of interest. Each event is represented as a set of values of monitored parameters. Consider a sensor network of n sensors $s_1, s_2, ..., s_n$ connected to a base station, where $x_i^{(t)}$ denotes a measured sensor value of a monitored parameter x_i received at a time slot t from sensor s_i. Then, n-tuple $\langle x_i^{(t)}, x_2^{(t)}, ..., x_n^{(t)} \rangle$ is an event e_t that represents states of monitored sensors at a time slot t. The sensor monitoring system analyzes sequences of events $e_1, e_2, \ldots, e_n, \ldots$ with the main goal to detect subsequences of events that may indicate some predefine activities of interest. Examples

of such activities include fire development in some areas of the monitored building or patient arrhythmias detected by a wireless health monitoring system.

3.3 Decomposing Status Information: Individual Shares and Parameter Dependencies

Our approach is based on secure multi-party computation where each sensor node delivers a part of the sensed data, called a share. Each sensor share is a subset of monitored parameters assigned to that sensor, for example can define a function *Share* that maps individual sensors in a power set of monitored parameters:

Share: ***Sensors*** $\rightarrow$ ***P(Parameters)***

Thus, in order to obtain complete information about the monitored environment (status information) a base station should collect shares from all sensors in the network. The shares should be selected in such a way that individual sensor outputs are unintelligible for reconstructing complete status information. Intelligible reconstruction of the status information is only possible when a certain number N of distinct shares is available. We will call N intelligibility threshold. In the following consideration we will associate each share with a monitored parameter. Thus for a sensor network with k nodes a complete status information includes knowledge of all sensor parameters. The complete status information is associated with the lowest security and privacy requirements, since it reveals all the data delivered by sensors.

In addition, we specify a set of parameter dependencies $xi, \ldots, xj \rightarrow xm, \ldots, xn$ reflecting the fact that from knowledge of parameters $xi, \ldots, xj$ we can (uniquely) infer the knowledge of parameters $xm, \ldots, xn$. The basic set of parameter dependencies comes from application semantics[2)]. In addition to the basic set, we can infer derived dependencies using the following Armstrong axioms [SKS05]:

- *Reflexivity:* $X \rightarrow Y$ *iff* $Y \subseteq X$;
- *Transitivity:* $X \rightarrow Y$, $Y \rightarrow Z$ *implies* $X \rightarrow Z$;
- *Augmentation:* $X \rightarrow Z$ *implies* $XY \rightarrow ZY$ *for any* Y.

We define a closure of a parameter X, denoted $X+$, as a set of all parameters Y such that parameter dependency $X \rightarrow Y$ can be inferred from a basic set of parameter dependencies using Armstrong axioms. If at least $n < k$ distinct shares (parameters) are required that allow base station to reconstruct complete status, then we say that the intelligibility threshold for this security level is n.

Example 1. Consider the following complete set of patient parameters: *patient_walking_speed, patient_pulse, patient_age*. The complete status information will include knowledge of all above parameters, thus $k = 3$. Consider also the following basic parameter dependency:

patient_walking_speed, patient_pulse $\rightarrow$ *patient_age*

Using Armstrong axioms we can infer all parameters from *patient_walking_speed* and *patient_pulse*. Thus, in this case the intelligibility threshold $n = 2$. Note that here we do not have a parameter dependency *patient_id* $\rightarrow$ *patient_age*, which would be natural to assume in the relational databases. The reason is that our dependencies reflect domain constraints, which are not necessarily stated explicitly as tuples in relational tables. Instead, they specify that there is an insecure way to use the knowledge of the left side of the dependency to infer the knowledge of the right side. Although in general it is natural to assume *patient_id* $\rightarrow$ *patient_age*, the association between patients' identities and ages can be stored in a highly secured database within a medical facility. Thus, there is no insecure way to receive *patient_age* from *patient_id*. However, observing *patient_walking_speed*, and *patient_pulse* an intruder can figure out the patient age with high certainty.

Thus, a part of the complete status information that can be received from information delivered by a sensor $si \in$ *Sensors* can be defined as a closure *Share(si)+* of the parameters assigned to *si*. A measure of certainty in the range 0-1 can be associated with each dependency and integrated in the inference process. Consideration of uncertain inference is out of scope of this paper.

3.4 Hiding Status Information: Refinement Hierarchy

In general, the intelligibility threshold N differs with changes in security and privacy requirements. For higher security, where sensor networks are allowed to reveal complete status information, the value of N is lower. In addition to share decomposition, we suggest ranking the security and privacy requirements using a refinement hierarchy. The ranking is based on logical specification of the data which is appropriate to reveal under given security constraints. For example, consider a body temperature sensor. The output of the sensor can correspond to one of the statements $P1$-$P3$ shown in Table 1 together with corresponding refinement levels. In this case statement $P3$ corresponds to the lowest security requirements since it reveals the

2) *In this sense parameter dependencies are similar to functional dependencies in relational databases.*

Refinement Level	Sensor Value	
0	temp is high	(*P*1)
1	temp > 38	(*P*2)
2	temp = 38.5	(*P*3)

Table 1 Refining sensor readings

most accurate temperature value and thus refines both *P*2 and *P*3, while *P*3 communicates the least amount of data and so it corresponds to the highest security requirements out of *P*1, *P*2 and *P*3.

In general the refinement hierarchy corresponds to Information Hiding (IH) rules that define how more refined sensor readings are mapped in less refined ones. Table 2 illustrates IH rules for sensors reporting patient's speed and pulse. The IH rules maps actual values of the speed and pulse into less refined ranges.

In the next subsection we will illustrate how refinement and share decomposition can be combined using dependency rules in a general framework.

3.5 Combining Share Decomposition and Refinement Using Data Dependencies

In this section we will extend axioms for dependency-based reasoning to capture both share decomposition and refinement hierarchy in the concept of node refinement. Informally, sensor node *N*1 refines a node *N*2 if knowledge of information transmitted by *N*1 can give us the knowledge of information transmitted by *N*2. Thus, the more refined node *N*1 provides less information protection than less refined node *N*2. The node refinement concept will be used to characterize information routing paths in sensor networks with respect to the information protection characteristics.

We will represent the integrated framework as a formal theory [EH90] consisting of the basic signature, dependency-based inference rules, and formal definition of the node refinement concept. The specification of a basic **Dep** signature is presented in Figure 3. It includes a set of sorts together with operations defined on them. The signature includes two user-defined sorts *Node* and *Parameter*. Five operations of the signature are as follows: *share(N)* associates a node *N* with a set of parameters that constitute the node's share (*P(Parameter)* denotes the power set of parameters); *refLevel(N,P)* specifies refinement level of a parameter *P* associated with a node *N*; *dep(P*1,*P*2*)* operation specifies that there is a semantic dependency between parameters *P*1 and *P*2, namely knowledge of *P*1 will provide us with knowledge of *P*2; *closure(N)* associates node *N* with a set of parameters that can be inferred from *share(N)* using dependency inference axioms; *refines(N*1,*N*2*)* is a predicate which is true if parameters in *closure(N*1*)* refine parameters in *closure(N*1*)*.

Dependency inference rules are presented in Figure 4. An inference rule *Premise* → *Conclusion* reflects the fact that there is a one step inference from *Premise* to *Conclusion*. Rules 1-3 correspond to basic Armstrong axioms.

Using the dependency axioms we define *closure* and *refines* operations in Figure 5.

Axiom 2 in Figure 5 reflects the fact that refinement introduces a special kind of parameter dependencies

Information Hiding (IH) rules	Less refined sensor readings confirming with the IH rules		Refined sensor readings	
Speed:	*Speed*	*Pulse*	*Speed*	*Pulse*
Slow: $S < 3$ mph	Slow	Normal	2	64
Moderate: 3 mph ≥ S ≥ 10 mph	Slow	Normal	2	69
Fast: 10 mph ≥ S	Slow	Normal	3	78
	Moderate	Normal	10	84
Pulse:	Fast	High	12	137
Low: $P > 55$ bpm	Fast	High	12	146
Normal: 55 bpm ≥ $P > 100$ bpm	Fast	High	10	144
High: 100 bpm ≥ $P > 180$ bpm	Moderate	High	3	146
Very high: $P \geq 180$ bpm	...	...	...	...

Table 2 Refinement and Information Hiding Rules

Signature: **Dep**

Sorts: ***Node, Parameter***

Operations:

share : ***Node*** → ***P*(*Parameter*)**

refLevel : ***Node, Parameter*** → ***Integer***

dep : ***P*(*Parameter*), *P*(*Parameter*)** → ***Boolean***

closure : ***Node*** → ***P*(*Parameter*)**

refines : ***Node, Node*** → ***Boolean***

Figure 3 Basic dep signature

on different refinement levels. A more refined version of a parameter implies a less refined one. Consider the example in Table 1 from the previous subsection. Knowledge of a more refined version of the temp parameter (e.g. *temp* = 38.5) gives us knowledge of a less refined version (e.g. *temp* > 38.5). We will denote it $temp_2 \rightarrow temp_1$, where indices correspond to predefined refinement levels.

We will also use the concept of node refinement with respect to a given parameter. The specification of the corresponding operation is as follows:

par-refines: ***Node***, ***Node***, ***Parameter*** → ***Boolean***
$(\forall N1,N2 \in Node, \forall X \in Parameter: X \in closure(N2) \cap closure(N1)$
$\wedge\ refLevel(N2,X) \leq refLevel(N1,X) \leftrightarrow$
par-refines$(N1,N2,X))$

Next, we will apply the introduced formalism to define the concept of privacy-aware routing in WSNs.

3.6 Privacy-aware routing

In the previous sub-sections we explained how the concepts of share decomposition, parameter dependencies and node refinement can be used to control information hiding. In this sub-section we will extend the proposed formalism to perform privacy-aware information routing in WSNs. In order to do that we extend the basic specification with an extra sort called *Route*. Intuitively, values of *Route* are legal routes in a sensor network including sensor nodes participating in the routing. We will introduce the following operations to provide axiomatic characteristics of the routing:

- *src*: ***Route*** → ***Node***. This operation returns a source node of a given route.
- *dest*: ***Route*** → ***Node***. This operation returns a destination node of a given route.
- *one_hop*: ***Node***, ***Node*** → ***Boolean***. This operation is a predicate that checks if there is a one-hop route between the two nodes.

We leave above three operations underspecified assuming that the relevant definitions should involve lower level network topology concepts.

- *route*: ***Node***, ***Node*** → ***Boolean***. This operation is a predicate that checks if there is a route between the two nodes. This predicate is defined as follows:

$(\forall Src, Dest \in Node: route(Src, Dest) \leftrightarrow$
$(one\_hop(Src, Dest) \vee \exists Z \in Node$
$(one\_hop(Src, Z) \wedge route(Z, Dest))))$

Above operations allow us to specify nodes belonging to a route:

1. Dependency reflexivity	$(\forall X,Y \in P(Parameter): Y \subseteq X \rightarrow dep(X,Y))$
2. Dependency transitivity	$(\forall X,Y \in P(Parameter): (\exists Z \in P(Parameter): dep(X,Z), dep(Z,Y)) \rightarrow dep(X,Y))$
3. Dependency augmentation	$(\forall X,Y,Z \in P(Parameter): dep(X,Y) \rightarrow dep(X \cup Z, Y \cup Z))$

Figure 4 Basic dependency inference rules

1. Parameter closure:
$(\forall N \in Node, \forall X,Y \in P(Parameter): X \subseteq share(N) \vee (X \subseteq closure(N) \wedge dep(X,Y)) \leftrightarrow Y \subseteq closure(N)$

2. Node refinement:
$(\forall N1,N2 \in Node: closure(N2) \subseteq closure(N1) \wedge$
$(\forall X \in Parameter: X \in closure(N2) \rightarrow refLevel(N2,X) \leq refLevel(N1,X)) \leftrightarrow refines(N1,N2))$

Figure 5 Definition of parameter closure and node refinement

$(\forall R \in Route, \forall N \in Node: N \in R \leftrightarrow$
$N = src(R) \vee N = dest(R) \vee (route(src(R), N) \wedge$
$route(N, dest(R))$

Next we introduce the concept of route refinement. Intuitively, route $R1$ refines a route $R2$ with respect to a parameter P if the least refined node of $R1$ refines the least refined node of $R2$ with respect to P. A definition of the least refined node in the route with respect to a parameter is as follows:

- *least-refined*: ***Route***, ***Node***, ***Parameter*** $\rightarrow$ ***Boolean***
$(\forall R \in Route, \forall X \in Parameter, \forall N \in Node:$
$N \in R \rightarrow (\forall N1,N2 \in Node: N1 \in R \rightarrow$
$par\text{-}refines(N1,N2,X)) \leftrightarrow least\text{-}refined(R, N, X))$

Now we are ready to define a key concept of refinement with respect to a given parameter. The definition is as follows:

route_par_refine: ***Route***, ***Route***, ***Parameter*** $\rightarrow$ ***Boolean***
$(\forall R1,R2 \in Route, \forall N1,N2 \in Node, \forall X \in Parameter:$
$(least\_refined(R1, N1, X) \wedge least\_refined(R2, N2, X)$
$\rightarrow par\text{-}refines(N1,N2,X)) \leftrightarrow$
$route\_par\_refine(R1,R2,X))$

Using the route refinement operation we can define different privacy preserving routing alternatives. The idea is to use alternative routes in order to tune privacy of the passing information. Consider two routes $R1$ and $R2$ delivering some parameter X in a wireless sensor network to a destination node (for example base station). Let us assume that we consider $R1$ as a main route and the system needs to evaluate privacy preservation options while substituting $R1$ with an alternative route $R2$. The alternative route can either preserve the same privacy level, reveal some additional information, or hide some information comparing to the data sent over the original route. We can use the introduced formalism to express those options as follows:

- Information-preserving alternative route selection:
$route\_par\_refine(R1,R2,X) \wedge$
$route\_par\_refine(R2,R1,X)$

- Information-revealing alternative route selection:
$route\_par\_refine(R2,R1,X)$

- Information-hiding alternative route selection:
$route\_par\_refine(R1,R2,X)$

In the next section we will apply the introduced formalism for the task of secure selection of alternative transmission schedules in WSNs. We will illustrate how our formalism that reflects application level data constraints in the form of data dependencies can be combined with lower network layer characteristics, such as collision handling.

4 Case Study: Integrating Dependency Formalism with Collision-Aware Transmission Scheduling

As we already mentioned in the introduction, one motivation for our research is to develop an integrated framework that would facilitate interaction between higher (logical) level security and privacy policies, and low-level security primitives based on physical characteristics of WSNs. Such an integrated framework aims to achieve the highest efficiency for a desired degree of security. Data dependency formalism presented in Section 3 allows us to characterize privacy and security requirements at application semantics (logical) level. In this section we will illustrate how this formalism can be integrated with lower level features of WSNs. As a case study we consider the problem of collision-aware transmission scheduling that takes into account information about collision domains of wireless transmissions.

4.1 Collision-Aware Transmission Scheduling

In general, the transmissions between sensors are *ad hoc* dependent on the query and require the use of a medium access control (MAC) layer to handle transmissions on the same medium. If we assume that all sensor nodes use the same frequency band for transmission, two transmissions that overlap will get corrupted (collide) if the sensor nodes involved in transmission or reception are in the same *collision domain* $CD(ni,nj)$ defined as the union of the transmission ranges of ni and nj. Figure 6 elaborates on the concept of collision domains in a typical wireless network such as IEEE 802.15.4 [ZL04] and illustrates how collisions are handled in such a network. Consider two sensor nodes $n1$ and $n2$ that wish to communicate. In Figure 6, sensor nodes $n1$, $n2$, $n3$, and $n4$, $n5$ and $n6$ are in the collision domain $CD(n1,n2)$. This implies that when $n1$ and $n2$ are communicating, $n3$, $n4$, $n5$ and $n6$ cannot participate in any communication. A typical wireless network handles collisions using carrier sense multiple-access with collision avoidance (CSCMA-CA) [Cro97]. In general, before starting a transmission, nodes must sense the channel for a predetermined amount of time (waiting time). If the channel is busy, the nodes wait for the predetermined amount of time after the channel becomes free. In addition, nodes back off for a random time to avoid the possibility that two or more nodes transmit at the same time after the waiting period. For this entire period, the node must sense the channel and this consumes energy. Each packet

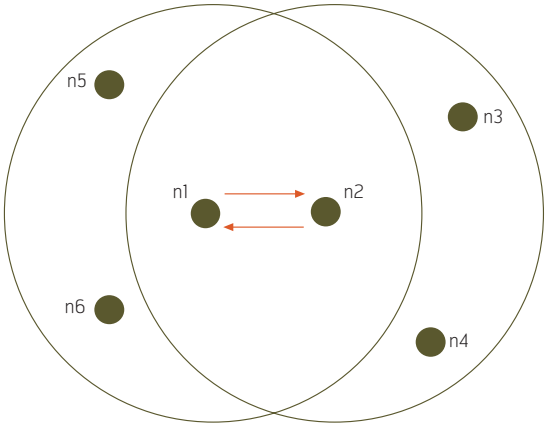


Figure 6 Collision domain of two communicating sensors

also needs to be acknowledged by the receiver since wireless channels are unreliable.

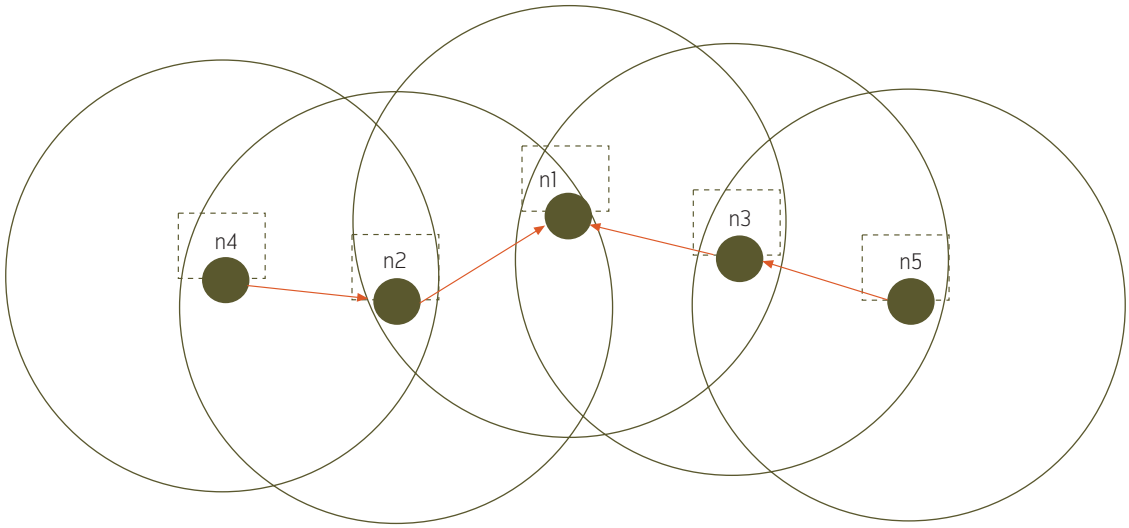
In related research [ZCK04, ZSKL05] we proposed a cost based algebraic query optimization framework for sensor networks. Using a Data Transmission Algebra (DTA) our framework generates an algebraic transmission schedule and assigns a time and energy cost to this schedule. The DTA consists of a set of operations that take transmissions between wireless sensor nodes as input and produce a schedule of transmissions as their result. A one-hop transmission from a source sensor node *ni* to a destination node *nj* is called elementary transmission (denoted *ni~nj*). Each elementary transmission *ni~nj* is associated with a collision domain *CD*(*ni,nj*) defined as a union of transmission ranges of *ni* and *nj*. A transmission schedule is either an elementary transmission, or a composition of elementary transmissions using one of the operations of the DTA. The basic DTA includes three operations that combine two transmission schedules A and B:

1 ***o*(A,B)**. This is a strict order operation; that is, A must be executed before B.

2 ***c*(A,B)**. This is a non-strict order operation; that is, either A executes before B, or vice versa. Thus, *c*(A,B) ≡(*o*(A,B) or *o*(B,A)).

3 *a*(**A,B**). This is an overlap operation; that is, A and B can be executed concurrently.

For an example of the DTA operations consider the query tree in Figure 7 which was generated for some query Q. It shows some DTA specifications that reflect basic constraints of the query tree. For instance, operation *o*(*n*4~*n*2, *n*2~*n*1) specifies that transmission *n*2~*n*1 occurs after *n*4~*n*2 is completed. This constraint reflects a part of the query tree topology. Operation *c*(*n*2~*n*1, *n*3~*n*1) specifies that there



Elementary transmission:	n4 ~ n2
Strictly ordered transmissions:	o(n4 ~ n2, n2~n1)
Non strictly ordered transmissions:	c(n2 ~ n1, n3~n1)
Concurrent transmissions:	a(n4 ~ n2, n5~n3)
Complete schedule:	o(a(n4 ~ n2, n5~n3), c(n2 ~ n1, n3~n1))

Initial specification:
n4~n2
n2~n1
n5~n3
n3~n1

o(n4 ~ n2,n2~n1)
o(n5 ~ n3, n3~n1)

c(n2 ~ n1, n3~n1)

a(n4 ~ n2, n5~n3)
a(n4 ~ n2, n3~n1)
a(n5 ~ n3, n2~n1)

Figure 7 Example of a query tree for some query Q and DTA specification

is an order between transmissions $n2{\sim}n1$ and $n3{\sim}n1$ since they share the same destination. However this order is not strict. Operation $a(n4{\sim}n2, n5{\sim}n3)$ specifies that $n4{\sim}n2$ can be executed concurrently with $n5{\sim}n3$, since neither $n3$ nor $n5$ belong to $CD(n4,n2)$, and neither $n4$ nor $n2$ are in $CD(n5,n3)$.

Each operation of the DTA specification defines a transmission schedule. The DTA introduces a set of transformation rules [ZCK04, ZSKL05] that can be used to generate more complex schedules. Figure 7 shows an example of a complete schedule that includes all elementary transmissions of the query tree. Figure 7 also shows the initial DTA specification reflecting basic constraints of the query tree. The initial specification consists of a set of elementary transmissions reflecting the tree topology imposed by the query semantics, as well as order and overlap operations over the elementary transmissions. Non-strict order constraints can be derived from the initial specification.

4.2 Security-Aware Transmission Scheduling in WSNs

In this section we will integrate the dependency based privacy preservation formalism with collision-aware transmission scheduling. First we specify a DTA theory consisting of the DTA signature and DTA inference rules. After that we will merge the Dep and DTA specifications in an integrated theory. Finally, we will extend the DTA cost model to assess privacy level for transmission schedules.

Signature: **DTA**

Sorts: ***Node, Schedule***

Operations:

$\sim$: ***Node, Node*** $\rightarrow$ ***Schedule***

o : ***Schedule, Schedule*** $\rightarrow$ ***Schedule***

c : ***Schedule, Schedule*** $\rightarrow$ ***Schedule***

a : ***Schedule, Schedule*** $\rightarrow$ ***Schedule***

Figure 8 DTA Signature

1. Order introduction	$N1 \sim N2, N2 \sim N3 \rightarrow o(N1 \sim N2, N2 \sim N3)$
2. Order transitivity	$o(X,Z), o(Z,Y) \rightarrow o(X,Y)$
3. Choice commutativity	$c(X,Y) \leftrightarrow c(Y,X)$
4. Overlap commutativity	$a(X,Y) \leftrightarrow a(Y,X)$
5. Left sub-schedule order	$(\exists S_i \in subs(S), o(X, S_i)) \rightarrow o(X,S)$
6. Right sub-schedule order	$(\exists S_i \in subs(S), o(S_i, X)) \rightarrow o(S,X)$
7. Sub-schedule choice	$(\exists S_i \in subs(S), c(X, S_i)) \rightarrow c(X,S)$
8. Sub-schedule overlap	$(\forall S_i \in subs(S), a(X, S_i)) \rightarrow a(X,S)$

Figure 9 Basic DTA Inference Rules

The DTA signature specification is presented in Figure 8. It includes a set of sorts together with operations defined on them. The DTA signature includes two sorts *Node* and *Schedule*. DTA schedules are explained informally in Section 4.1. Elementary transmission (denoted ~) is a DTA operation that takes two nodes as input and outputs a schedule. The rest of the DTA operations (o, c, a) map two input schedules to an output schedule.

In order to introduce DTA axioms that define the above operations we extend the basic DTA signature with a secondary operation subs, which for a given DTA schedule returns all its sub-schedules. The subs operation is specified as follows:

$$subs : \textbf{\textit{Schedule}} \rightarrow \textbf{\textit{P}}(\textbf{\textit{Schedule}}),$$

where $\textbf{\textit{P}}(\textbf{\textit{Schedule}})$ denotes the power set of schedules. The following equations complete the specification of subs:

$$subs(X{\sim}Y) = \{X{\sim}Y\}.$$
$$subs(comp(S1,S2)) = \{comp(S1,S2)\} \bigcup subs(S1) \bigcup subs(S2),$$

where *comp* denotes any of DTA operations o, c, or a.

DTA axioms are represented in Figure 9 in the form of inference rules. For example, using rule 1 (*order introduction*) we can infer a strict order of two elementary transmissions if the destination node of the first transmission is also a source node of the second transmission. Rule 5 generates a strict order of DTA schedules X and S if there exists a sub-schedule S_i of the schedule S such that $o(X,S_i)$ can be generated by the DTA rules. In order to infer $a(X,S)$, we should be able to infer $a(X, S_i)$ for all sub-schedules S_i of the schedule S.

We link together **Dep** and **DTA** specifications with the following scheduled routes operations:

$$scheduledRoutes: \textbf{\textit{Schedule}} \rightarrow \textbf{\textit{P}}(\textbf{\textit{Route}})$$

This operation associates a set of valid routes with a DTA schedule. Note that a DTA schedule may include one or more routes whose transmissions can be scheduled either serially or concurrently. The following equation provides specification of *scheduledRoutes* for elementary DTA transmission:

$scheduledRoutes(X\sim Y) = \{R\} \leftrightarrow src(R) = X \wedge dest(R) = Y$

Routes for more complex DTA schedules can be generated by recursive application of the *scheduledRoutes* over all elementary transmissions of complex schedules.

We propose to extend the DTA cost model using the decomposition measures of share-base data delivery. In this case, in addition to time and energy, each transmission schedule can be associated with certain security requirements. The optimizer should choose a schedule with the best response time, energy consumption and acceptable security value. Below we consider only the time cost measure that can be specified as the following operation:

$TimeCost$: ***Schedule*** $\rightarrow$ ***Real***

Figure 10 shows time cost estimation expressions for each of the DTA expressions. In this case, the cost corresponds to the execution time associated with a particular schedule. For clarity of presentation we ignore energy consumption at this point. For example, the execution time of elementary transmission *ni~nj* consists of local processing times *Tp* at nodes *ni* and *nj* plus the time *Ttx* required for transmitting data from *ni* to *nj*. The execution time of strict order of schedules A and B is the sum of execution times of A and B. For overlapping schedules A and B, the execution time would be the maximum of the execution times of A and B. Finally, execution time of the choice between A and B is the same as the execution time of the strict order minus a predefined time factor *Tf*. *Tf* is a non-negative number indicating that in general, the optimizer prefers the choice operation over strict order, since the latter restricts flexibility of the optimizer in query scheduling. We ignore propagation times as they are negligible in this case.

Next we define a numeric security metric associated with each route. To do that we will introduce the following operations:

- *weight*: ***Parameter*** $\rightarrow$ ***Real***. This operation assigns an importance weight to each parameter delivered by WSN.
- *maxRefLevel*: ***Parameter*** $\rightarrow$ ***Integer***. This operation assigns the maximum refinement level to each parameter delivered by WSN.
- *refRatio*: ***Parameter***, ***Node*** $\rightarrow$ ***Real***. This operation evaluates a parameter refinement ratio associated with wireless node. The definition of this operation is as follows:

Schedule	TimeCost
ni~nj	*Tp*(*ni*) + *Ttx*(*ni~nj*) + *Tp*(*nj*)
o(*A*,*B*)	*TimeCost*(*A*) + *TimeCost*(*B*)
a(*A*,*B*)	*max*(*TimeCost*(*A*), *TimeCost*(*B*))
c(*A*,*B*)	*TimeCost*(*A*) + *TimeCost*(*B*) − *Tf*

Figure 10 Estimating time costs of transmission schedules

$(\forall X \in Parameter, \forall N \in Node: refRatio(X,N) = refLevel(X,N) / maxRefLevel(X,N)$.

- *leastRfNode*: ***Route***, ***Parameter*** $\rightarrow$ ***Node***. This operation returns a least refined node for given parameter refinement in a route. The definition of this operation is as follows:

$(\forall R \in Route, \forall X \in Parameter, \forall N \in Node: least\text{-}refined(R,N,X) \leftrightarrow N = leastRfNode(R,X))$

Now we are ready to introduce the *SecurityMeasure* operation that evaluates the measure of security for a route in WSN. Informally, the measure of security for a route *R* is computed as a weighted average of refinement ratios of each parameter associated with a least refined node for that parameter within the route *R*. Here is a formal definition of *SecurityMeasure* operation:

$SecurityMeasure$: ***Route*** $\rightarrow$ ***Real***
$(\forall R \in Route: SecurityMeasure(R) =$

$$\sum_{P_i} weight(P_i) \times refRatio(P_i, leastRfNode(R,P_i)))$$

Consider an example of WSN delivering three parameters *P*1, *P*2 and *P*3 and two routes *R*1 and *R*2 with the weights and refinement ratio distributions shown in Figure 11. For each route Figure 11 shows parameter name, parameter weight and refinement ratio for least refined nodes in corresponding routes with respect to given parameter. The security costs of each route will be evaluated as follows:

$SecurityMeasure(R1) = 1 * 0.5 + 2 * 0.8 + 1 * 0 = 2.1;$

$SecurityMeasure(R2) = 1 * 0.7 + 2 * 0.5 + 1 * 0.1 = 1.8.$

Thus, route *R*1 reveals more information and should be considered as less secure compared to route *R*2.

Finally, we define an integrated time/security measure that combines both time cost and security measure. We specify the *TimeSecurityMeasure* operation that for each DTA schedule returns a pair consisting

R1	Parameter	P1	P2	P3	R2	Parameter	P1	P2	P3
	Weight	1	2	1		Weight	1	2	1
	Refinement Ratio	0.5	0.8	0		Refinement Ratio	0.7	0.5	0.1

Figure 11 Example of parameter distributions for routes R1 and R2

of the time cost of the schedule and the minimal security measure out of all the routes associated with the DTA schedule:

$$TimeSecurityMeasure: \mathbf{Schedule} \rightarrow \mathbf{Real} \times \mathbf{Real}$$
$$(\forall S \in Schedule: TimeSecurityMeasure(S) = (V1,V2) \leftrightarrow$$
$$V1 = TimeCost(S)\ V2 = min(\{SMi \mid SMi = SecurityMeasure(Ri) \text{ and } Ri \in scheduledRoutes(S)\}$$

Note that TimeSecurityMeasure estimates the least secure route in the DTA schedule. We could also tune the TimeSecurityMeasure definition for the most secure route, or for an average security value. Similarly, we can define a more aggregated measure that in addition to time cost and security measure would take into account other important characteristics of WSN such as energy consumption and quality of data. The optimizer will have to implement flexible multiobjective optimization strategies [Miett99] using utility functions assigning relative importance to each of the above metrics. This subject is associated with important and interesting research issues, which are out of scope of this paper.

5 Conclusion

We have considered a logical framework that utilizes taxonomy of the security and privacy solutions based on share decomposition and parameter refinement. The proposed approach provides a systematic and efficient way for privacy preserving data management, secure data distribution, integration and searching that compliment existing methods and techniques implemented on lower network layers. The proposed approach can be considered as a step towards fulfilling the need for infrastructures integrating security and privacy policies, security protocols and low-level security primitives to achieve the highest efficiency for a desired degree of security. In particular, we demonstrated how our approach can be integrated with a lower level optimization framework that performs collision aware transmission scheduling. Further work should be focused on efficient implementation of the proposed framework. This includes implementing flexible multiobjective optimization strategies that would provide secure and efficient data delivery in data intensive wireless sensor networks.

Acknowledgement

We would like to thank Rune Fensli for his most valuable feedback on this paper and for his considerable help with motivating scenarios for this research.

References

[Ana05] Anand, M, Ives, Z, Lee, I. Quantifying Eavesdropping Vulnerability in Sensor Networks. *Proc. DMSN'05*, August 2005.

[Arg05] Argyroudis, P, O'Mahony, D. Secure Routing for Mobile Ad Hoc Networks. *IEEE Communication Surveys*, Third Quarter, 2005.

[Bru05] Bruck, J, Gao, J, Jiang, A. MAP: Medial Axis Based Geometric Routing in Sensor Networks. *Mobicom'05*, Cologne, Germany 2005.

[Buc02] Buchegger, S, Le Boudec, J-Y. Performance Analysis of the CONFIDANT Protocol. *Proc. Mobihoc'02*, 226-236, 2002.

[Cha03] Chan, H, Perrig, A, Song, D. Random key predistribution schemes for sensor networks. *IEEE Symposium on Security and Privacy*, 197-213, May 2003.

[Cro97] Crow, B, Widjaja, I, Kim, L G, Sakai, P T. IEEE 802.11 Wireless Local Area Networks. *IEEE Communications Magazine*, 35 (9), 116-126, 1997.

[Dem06] Demirbas, M, Song, Y. An RSSI-based Scheme for Sybil Attack Detection in Wireless Sensor Networks. *Proceedings of the 2006 international Symposium on on World of Wireless, Mobile and Multimedia Networks*, 564-570, 26-29 June 2006.

[Den03] Deng, J, Han R, Mishra, S. Security Support for In-network Processing in Wireless Sensor Networks. *Proc. ACM SASN*, 2003.

[Din06] Dini, G, Savino, I M. An Efficient Key Revocation Protocol for Wireless Sensor Networks. *Proc. 2006 International Symposium on on World of Wireless, Mobile and Multimedia Networks*, 450-452, 2006.

[Djen05] Djenouri, D, Khelladi, L, Badache, N. A Survey of Security Issues in Mobile Ad Hoc and Sensor Networks. *IEEE Communications Surveys*, 7 (4), Fourth Quarter, 2005.

[Du03] Du, W et al. A Pairwise Key Pre-distribution Scheme for Wireless Sensor Networks. *Proceedings of the 10th ACM Conference on Computer and Communications Security*, October 2003.

[EH90] Ehrig, H, Mahr, B. *Fundamentals of Algebraic Specification. Equations and Initial Semantics*. Springer-Verlag, 1990.

[Esc02] Eschenauer, L, Gligor, V D. A key-management scheme for distributed sensor networks. *Proceedings of the 9th ACM Conference on Computer and Communications Security*, November 2002.

[Gan03] Ganesan, P et al. Analyzing and Modeling Encryption Overhead for Sensor Network Nodes. *Proc. WSNA'03*, September 2003.

[Gan05] Ganeriwal, S et al. Secure Time Synchronization Service for Sensor Networks. *Proc. ACM WiSE*, September 2005.

[Heg06] Hegland, A M et al. A Survey of Key Management in Ad Hoc Networks. *IEEE Communications Surveys*, 8 (3), Third Quarter, 2006.

[Law06] Law, Y W, Doumen, J, Hartel, P. Survey and Benchmark of Block Ciphers for Wireless Sensore Networks. *ACM Transactions on Sensor Networks*, 2 (1), 65-93, 2006.

[Laz05] Lazos, L, Poovendran, R. SeRLoc: Robust Localization for Wireless Sensor Networks. *ACM Transactions on Sensor Networks (TOSN)*, 1 (1), 2005.

[Man05] Manzo, M, Roosta, T, Sastry, S. Time Synchronization Attacks in Sensor Networks. *Proc. ACM SASN*, November 2005.

[Mar00] Marti, S et al. Mitigating Routing Misbehavior in Mobile Ad Hoc Networks. *Proc. Mobicom'00*, 255-265, August 2000.

[Miett99] Miettinen, K. *Nonlinear Multiobjective Optimization*. Kluwer Academic Publisher, 1999.

[New04] Newsome, J et al. The Sybil Attack in Sensor Networks: Analysis & Defenses. *Proc. ACM IPSN'04*, April 2004.

[Pras04] Prasithsangaree, P, Krishnamurthy, P. On a Framework for Energy-Efficient Security Protocols in Wireless Networks. *Computer Communications* (Special issue: Security and Performance in Wireless and Mobile Networks), 27, 1716-1729, 2004.

[Shi04] Shi, E, Perrig, A. Designing Secure Sensor Networks. *IEEE Wireless Communications*, 38-43, December 2004.

[SKS05] Silberschatz, A, Korth, H, Sudarshan, S. *Database System Concepts*, 5th edition. McGraw-Hill, 2005.

[Siv04] Sivrikaya, F, Yener, B. Time Synchronization in Sensor Networks: A Survey. *IEEE Network*, 18 (4), 45-50, 2004.

[Sun05] Sun, K, Ning, P, Wang, C. Fault-tolerant Cluster-wise Clock Synchronization for Wireless Sensor Networks. *IEEE Transactions of Dependable and Secure Computing*, 2 (3), 177-189, 2005.

[Wag04] Wagner, D. Resilient Aggregation in Sensor Networks. *Proc. ACM SASN*, October 2004.

[Wes06] Westoff, D et al. Concealed Data Aggregation for Reverse Multicast Traffic in Sensor Networks: Encryption, Key Distribution and Routing Adaptation. *IEEE Transactions on Mobile Computing*, 5 (10), 1417-1431, 2006.

[Woo02] Wood, A, Stankovic, J. Denial of Service in Sensor Networks. *IEEE Computer*, 35 (10), 54-62, 2002.

[Xu06] Xu, W, Ma, K, Trappe, W, Zhang, Y. Jamming Sensor Networks: Attack and Defense Strategies. *IEEE Network*, 41-47, May/June 2006.

[ZSKL05] Zadorozhny, V, Sharma, D, Krishnamurthy, P, Labrinidis, A. Tuning query performance in sensor databases. *Proc. of MDM*, 2005.

[ZCK04] Zadorozhny, V, Chrysanthis, P, Krishnamurthi, P. A Framework for Extending the Synergy between Query Optimization and MAC Layer in Sensor Networks. *Proceedings of the International Workshop on Data Management for Sensor Networks*. In conjunction with 30th International Conference on Very Large Data Bases, Toronto, Canada, 2004.

[Zha06] Zhang, L. A Self-adjusting Directed Random Walk Approach for Enhancing Source-Location Privacy in Sensor Network Routing. *Proc. IWCMC*, Vancouver, July 2006.

[Zhe06] Zheng, J, Lee, M J, Anshel, M. Toward Secure Low Rate Wireless Personal Area Networks. *IEEE Transactions on Mobile Computing*, 5 (10), 1361-1373, 2006.

[ZL04] Zheng, J, Lee, M. Will IEEE 802.15.4 Make Ubiquitous Networking a Reality? A Discussion on a Potential Low Power, Low Bit Rate Standard. *IEEE Communications Magazine*, June 2004.

For a presentation of Vladimir Zadorozhny, please turn to page 26.

For a presentation of Vladimir A. Oleshchuk, please turn to page 3.

Prashant Krishnamurthy is an Associate Professor with the Graduate Program in Telecommunications and Networking at the University of Pittsburgh, USA. He obtained his PhD in 1999 (when he also joined the University of Pittsburgh) from Worcester Polytechnic Institute, Massachusetts, USA. At Pitt, he regularly teaches courses on wireless communication systems and networks, cryptography and network security. His research interests are wireless network security, wireless data networks, position location in indoor wireless networks, and radio channel modeling for indoor wireless networks. He is the co-author of the book "Principles of Wireless Networks: A Unified Approach". He served as the chair of the IEEE Communications Society – Pittsburgh Chapter from 2000 to 2005.

email: prashant@tele.pitt.edu

RFID and Privacy

GEIR M. KØIEN

Geir M. Køien is a researcher in the Network Technologies group of Telenor R&I

In this short article we want to draw attention to privacy issues associated with the use of the RFID technology and to provide the interested reader with pointers to further reading.

What is RFID and What is it Used for?

An RFID (Radio Frequency Identification) device consists of a small and inexpensive microchip attached to an antenna. The device has very limited capabilities, both in terms of memory and processing capabilities. The chip can therefore be very small and some RFID tags are thin enough to be embedded in paper. Most RFID devices are only capable of transmitting a unique serial number a short distance and they do so in response to a request from a reading device. The RFID devices/tags come in many flavors and can broadly be divided into two main types: Passive (low cost) RFID tags and Active RFID devices.

The RFID technology is also being standardized. For instance, the *International Organization for Standardization (ISO)* has published several standards for RFID (including [1-4]). There are also industry consortia standards like the Electronic Product Code (EPC) standards [5].

Passive RFID Tags

Passive RFID tags have no internal power supply and they rely solely on the electrical current induced in the antenna by the incoming radio signal. This will provide enough power to transmit a response.

Since the passive tags do not need to include a battery the devices can be very small. One example is the Hitachi µ-chip. The µ-chip uses the frequency in the 2.45 GHz area. It has a 128-bit memory for storing an ID number. The physical size of the chip is a minute 0.4 mm^2. The µ-chip is small enough to be embedded in paper [6].

Figure 1 A passive Wal-Mart Electronic Product Code (EPC) RFID tag (Figure courtesy of Wikipedia)

However, the size of the RFID chips is somewhat misleading as a size indicator for passive tags since the external antenna tends to be in the order of 100 times bigger than the chip itself.

Low cost EPC RFID tags only cost around 5 cents (US) per tag. The EPC tags are a replacement for the bar codes commonly in use for inventory tracking today. When the antenna is included the tag will have the size of a postage stamp and upwards. Passive tags have practical read distances ranging from about 5-10 cm and up to a few meters depending on the radio frequency and size and design of the antenna.

Active RFID Devices

In contrast to passive RFID tags, the active RFID devices have their own internal power source (battery) to generate the response signal. Active devices are in general more reliable (fewer errors) than passive tags, but they also have a limited shelf life and are normally much more expensive than passive tags. The active devices can have much more memory and can perform substantial active processing. Since the active devices have their own power source they can also transmit at higher power levels than passive tags. This allows the active RFID devices to operate at longer ranges, up to 100 meters is not uncommon, and to operate in 'difficult' radio environments where passive tags would not have worked at all. Battery lifetimes between 5 and 10 years are not uncommon. The batteries are typically physically embedded (without any possibility of replacing it) in the device and the lifetime is decided by the lifetime of the battery. Active RFID devices come in many shapes and sizes, ranging from small and relatively inexpensive devices (coin sized, cost as little as 1-2 Euro) to expensive (>10-15 Euro) and relatively large devices.

RFID Usage Areas

RFID tags are already quite common in everyday life. Examples include:

ISSN 0085-7130

Figure 2 The AutoPASS tollroad active RFID device

- *Door key replacement*
  Physical proximity (5-50 cm) with the door is sufficient to unlock it.

- *Tickets*
  Tickets with embedded RFID chips are substantially harder to counterfeit and potentially provide much better admittance control.

- *Tagging of luggage (air travel)*
  Improvements in the airport baggage handling can only be a good thing. One thing not to worry about anymore[1)].

- *Books in libraries*
  The **Bibliotheca**[2)] BiblioChip® RFID library system would be just one example. There has been a fair amount of user privacy related controversy on tagging library books. A through technical discussion of the subject can be found in [7].

Figure 3 RFID implant[3)]

- *Automated road toll payment devices*
  In Norway a system called AutoPASS is used (see Figure 2). For more information see [8]. Similar systems are in operation in other countries.

- *Inventory tags*
  The EPC-type tags are now becoming ubiquitous and are used extensively in inventory management systems. See Figure 1 for an illustration of EPC tags.

- *Surgically embedded RFID tags for identification purposes*
  These are often used for cats and dogs, but some people have also voluntarily tagged themselves this way. (See Figures 3 and 4.) More information in [9,10].

- *License plates (cars)*
  Such plates are designed to be read at speeds of up to 300 km/h and up to 100 meters away. See the movie clip at [11] for, at least to the author, a chilling, visual demonstration of the UK e-plate initiative.

- *Id-cards, including passport*
  The use of RFID enabled identity cards is highly controversial, but the proliferation of RFID tags in passports etc seems only to increase.

A comprehensive list of usage areas can be found in [12-14].

RFID Trends

The cheap RFID tags are not very capable devices. Moore's law also applies to RFID devices and they are getting cheaper and cheaper to produce. In other areas this has meant a proliferation of more capable devices, but as it turns out this is not quite the case for RFID systems. Roughly speaking there are two main development trends, one for cheaper tags and one for more functionality in the devices.

The trend for more capable devices relies on active RFID devices. The active RFID devices provide some physical protection and are able to execute cryptographic operations. These devices are suitable for identity cards and key replacement access cards etc where credible security and privacy is important.

---

1) *The Monty Python song* "I'm so worried" *expresses sincere existential angst for the Heathrow baggage retrieval system (*"And I'm worried about the baggage retrieval system they've got at Heathrow" *[15]). This used to be a well-founded worry, but thanks to RFID tags this may now be a thing of the past!*

2) *The Bibliotheca homepage is at http://www.bibliotheca-rfid.com/*

3) *Picture courtesy of Amal Graafstra. Amal has two RFID implants, one in each hand. He can access his front door, car door, and log into his computer using his implants, and has written a book called* RFID Toys *[10], which details how to build these and other RFID enabled projects. See more information at http://amal.net/rfid.html*

Figure 4 X-ray of RFID implants[4)]

They are also well suited for usages that are relatively price insensitive, where the required security level is high and where the cost of distribution and deployment is high relative to the cost of the RFID device itself.

The other trend is pervasive inventory tracking. If you are to tag inexpensive goods then the RFID tag had better be cheap. The development of these tags is driven primarily towards cheaper and cheaper tags. Thus, these RFID tags will likely never have cryptographic processing or other features that can help with improving user privacy. That is, modern (inexpensive) RFID tags may implement a *kill command* that is used for destroying/disabling the device at checkout. The *kill command* is then activated by a unique/individual (per tag) code sequence [13]. There is no cryptographic protection as such of the *kill command*, just the assumption that the *code sequence* (32-bit code for EPC-type of tags) will not likely be triggered by accident. However, for the consumer it will in practice be impossible to know if the RFID tags are (all) killed.

The Threat to Personal Privacy from Pervasive RFID Deployment

Passive RFID Tags

We are now approaching a critical turning point at which inexpensive RFID tags will successively be used as a replacement for barcodes. RFID tags have two distinct characteristics that set them apart from traditional printed barcodes:

- *An RFID tag carries a unique identifier*
  The traditional barcode only indicates an object type. For example, a barcode printed on a box might state that the box contains chocolate bars and identify the manufacturer. An RFID tag, on the other hand, carries an extended serial number that not only identifies the chocolate bar type, but also identifies the individual chocolate bar. This permits fine-grained control over product distribution and permits tracking of individual inventory items.

- *An RFID tag may be read by radio contact*
  A barcode scanner must make close-range optical contact to read a barcode effectively. In contrast, an

4) *Picture courtesy of Amal Graafstra. You can see the larger 3 x 13 mm EM4102 chip in the left hand and the smaller 2 x 12 mm Philips HITAG S 2048 the right. More information is available at http://amal.net/rfid.html*

RFID tag may be read just by being placed in the vicinity of a reader. Indeed, a reader may be capable of scanning hundreds of RFID tags simultaneously. The practical reading distance depends on the size and design of the antenna, and generally the longer reaching devices require larger antennas or an active device.

The resolution of RFID identification is problematic seen from a personal privacy point of view. If you buy an RFID tagged chocolate bar and carry it with you then you can of course also be tracked. The tracker may or may not know your identity, but they can certainly track you if they have the appropriate reading device. The danger of being tracked is of course mitigated by the limited range of these tags and the use of *kill commands* could eliminate and/or reduce the risk[5)].

Active RFID Devices

The traditional active RFID devices are comparatively expensive and there are much fewer of them than for passive RFID tags. However, the active devices do have substantially longer range and are thus more exposed to illegitimate reading/eavesdropping.

The active RFID devices are also much more capable devices. They can store a lot more data and do a fair amount of active processing. The active RFID device may also contain cryptographic processing capabilities. Another characteristic of the active device is that it is typically embedded in items that the users keep with them (associated with them) for prolonged periods of time. For instance, RFID enabled identity cards would typically travel with the user. License plate tagging, as exemplified by the e-plate scheme [11] and the AutoPASS toll road payment scheme [8] are but two examples. Indeed, one should expect all valuable items to be tagged. Another example would be automated teller machines (ATM) that use RFID enabled ATM cards. There already exist microchip enabled ATM cards and making these cards communicate wirelessly is a logical next step[6)]. The idea has been patented in the US [16] and integration of RFID type of technology with mobile phones is already taking place [17].

There is therefore every possibility that we will carry a number of active RFID devices with us more or less continually. If these devices leak information to the surrounding environment then, we are susceptible to being tracked. This would constitute a clear threat to our personal identity- and location privacy. The threat can be mitigated by having a limited physical device reading range and by proper use of cryptographic methods to hide sensitive information.

Classification on Threats to Personal Privacy

In the overview article in IEEE Security and Privacy Magazine by Garfinkel, Juels and Pappu the authors outline a set of threats to personal privacy [13]. Keep in mind that a threat does only point to a certain general possibility; thus a threat is not the same as an actual attack. A research focused RFID privacy survey is found in [18]. The main threats as outlined in [13] are:

- *Action threat*
  This is a threat that is triggered if many tags are moved simultaneously in a shop. The fact that many tags are moved may indicate a case of shoplifting etc. This in turn may trigger (video) surveillance and direct action by the shop. The problem is of course that there may be many legitimate actions that can cause multiple tags to be moved simultaneously.

- *Association threat*
  The customer identity can be associated with the RFID Id-tag. Loyalty cards etc are in use in many shops. The EPC tag can then be linked with the loyalty card and a detailed database of customer spending patterns can be created. The threat is similar to a bar code association threat, but since the EPC codes have substantially higher resolution (identifying specific instances instead of type of object) the privacy threats worsen.

- *Location- and identity threats*
  Pervasive deployment of RFID technology means that it will be quite easy to place covert RFID tags such that one can track users. The tracking may be of unidentified users (called a *constellation threat* in [13]), but of course one may also identify the users and track specific users. It is noted that using multiple RFID sources may mean that a resourceful adversary can correlate information and gradually establish a user identity[7)]. The user will then lose his/her identity- and location privacy.

---

5) *The precision of the kill command cannot be guaranteed and, at least for now, many tags do not have the kill command feature.*

6) *The term Near Field Communication (NFC) is often used with these devices. NFC encompasses short-range wireless interaction in consumer electronics, mobile devices and PCs. More information on NFC can be found at http://www.nfc-forum.org/home, http://en.wikipedia.org/wiki/Near_Field_Communication and numerous other homepages. The GSM Association has produced a whitepaper on NFC available at [17].*

- *Preference threats*
  Obviously, the RFID technology can be used to establish a pattern of user behavior. This includes the preferences made by the user and extends beyond knowing what products you buy since it is also possible to identify each product individually. Thus, one can build up a history of your preferences which includes knowledge of when and where you bought certain items and even the price you paid for the products.

- *Transaction threats*
  If your whereabouts can be traced then one may also infer your activities. That is, one can derive information about who you meet and what you do. For instance, if it can be demonstrated that you were present at a certain location at a certain time then you would likely have met person X.

- *Breadcrumb threat*
  The breadcrumb threat arises from an association threat. If you are associated with a certain set of RFID tags/devices, then the signature of the set of RFID tags/devices de facto constitutes an identity. The problem is that it can be relatively easy to fake this implied/derived identity. Thus, if somebody gets access to that 'associated identity' then they can impersonate you. The threat arises from the illicit and invalid use of data, but this happens quite regularly and the threat is quite real.

As with all technology which may be used for identification purposes one also has the associated risk of wrongful identification. There can be many sources, but as cited in [13] many RFID devices have poor physical protection and they can easily be tampered with. Another real risk of derived (*constellation type* of) identification stems from poor data quality and accidental mistaken identification (either failure to identify the user at all or identifying the wrong user). The problem is that the system may not necessarily tell you that it has identified you, but act on the derived 'identity' nevertheless[8)].

The Use of RFID for Human Identification

This is and has been a highly controversial subject. For instance, in the US there is the example of the controversy surrounding the Department of Homeland Security (DHS) committee report *The Use of RFID for Human Identity Verification* [19]. The full committee advised caution in deployment of RFID for human identification purposes. However, in a subcommittee input document the advice was not to use RFID for human identification at all [20]. An account of the controversy appeared in [21].

The executive summary of the subcommittee input document explains very well the problems with RFID in passports etc. The following excerpt from the report nicely captures the situation [20]:

> "There appear to be specific, narrowly defined situations in which RFID is appropriate for human identification. Miners or firefighters might be appropriately identified using RFID because speed of identification is at a premium in dangerous situations and the need to verify the connection between a card and bearer is low.
>
> But for other applications related to human beings, RFID appears to offer little benefit when compared to the consequences it brings for privacy and data integrity. Instead, it increases risks to personal privacy and security, with no commensurate benefit for performance or national security. Most difficult and troubling is the situation in which RFID is ostensibly used for tracking objects (medicine containers, for example), but can in fact be used for monitoring human behavior. These types of uses are still being explored and remain difficult to predict.
>
> For these reasons, we recommend that RFID be disfavored for identifying and tracking human beings. When DHS does choose to use RFID to identify and track individuals, we recommend the implementation of the specific security and privacy safeguards described herein."

In California there was a proposal a few years ago for a scheme to provide school children with RFID tags. This was met with a public outcry and the scheme was promptly stopped. The California senate recently voted to ban RFID tagging of schoolkids. The bill, which expires in 2011, prohibits use of RFID in id-cards for schoolkids [22].

---

7) *A user can have many identities. An identity may also be considered an emergent property of a set of recognizable characteristics that may serve as a reference to the entity. Conceptually, an identity does not even have to be recognized by the identified entity(person) as long as other entities can use the identity as a reference to the target entity(person).*

8) *Even if the derived identity is correct it is still problematic if the identification process takes place without your knowledge or consent.*

Summary

In this article I have presented a brief introduction to the RFID technology in general and to some of the associated personal privacy problems. The article does not go into detail and is not comprehensive, but is rather aimed at providing a glimpse of an interesting topic. The RFID technology will affects us all and the author hopes the article has triggered the curiosity of the reader and he hopes the interested reader will pursue the topic further. The references should provide a reasonably good starting point.

References

1 ISO. *Identification cards – Contactless integrated circuit(s) cards – Vicinity cards – Part 1: Physical characteristics*. Geneva, Switzerland, 1-2007. (ISO/IEC 16693-1)

2 ISO. *Identification cards – Contactless integrated circuit(s) cards – Vicinity cards – Part 2: Air interface and initialization*. Geneva, Switzerland, 12-2006. (ISO/IEC 16693-2)

3 ISO. *Identification cards – Contactless integrated circuit(s) cards – Vicinity cards – Part 3: Anticollision and transmission protocol*. Geneva, Switzerland, 9-2005. (ISO/IEC 16693-3)

4 ISO. *Information technology – Radio frequency identification for item management – Part 1: Reference architecture and definition of parameters to be standardized*. Geneva, Switzerland, 4-2007. (ISO/IEC 18000-1) (The ISO/IEC 18000 consists of several parts. Only part 1 is included here.)

5 *Electronic Product Code Standards and Specifications*. Published by EPCglobal, available at http://www.epcglobalinc.org/standards/

6 *Hitachi Europe Limited, Mu Chip Data Sheet*. Available at http://www.hitachi-eu.com/mu/products/mu_chip_data_sheet.pdf

7 Molnar, D, Wagner, D. Privacy and Security in Library RFID: Issues, Practices, and Architectures. In: *Proceedings of the 11th ACM conference on Computer and Communications Security*, ACM Press, 210-219, 2004. (ISBN 1-58113-961-6)

8 *AutoPASS homepage* (in Norwegian). http://www.autopass.no/index.html

9 *Amal Graafstra, Amal's RFID implant page*. http://www.amal.net/rfid.html

10 Graafstra, A. *RFID Toys: 11 Cool Projects for Home, Office and Entertainment*. Wiley, 2006. (ISBN 13: 978-0471771968)

11 *E-Plate video*. Hills Numberplates Ltd, Birmingham, UK. Published at http://www.e-plate.com/videos/e-plate.mov. An e-plate brochure is also available at http://www.e-plate.com/e-Plate%20Brochure.pdf

12 Garfinkel, S, Rosenberg, B. *RFID: Applications, Security, and Privacy*. Addison-Wesley Professional, 2005. (ISBN 13: 978-0321290960)

13 Garfinkel, S, Juels, A, Pappu, R. RFID Privacy: An Overview of Problems and Proposed Solutions. *IEEE Security and Privacy Magazine*, 3 (3), 34-43, 2005.

14 *Wikipedia: Radio-frequency identification*, http://en.wikipedia.org/wiki/RFID. Note: Wikipedia articles may be edited at any time and so the contents and quality of this reference may change over time.

15 Monty Python. *Monty Python Sings*. Audio CD, Virgin Records, November 1991. The song *I'm So Worried* appears as track 10.

16 *Currency dispensing ATM with RFID card reader*. United States Patent 7004385, Registered to Douglass, Mark (North Canton, OH, US), 03/31/2004. Available at http://www.freepatentsonline.com/7004385.html

17 *GSMA Publishes White Paper On Near Field Communications (NFC)*. GSM Association Press Release 2007, Barcelona, Spain,13 February 2007. The press release is available at http://www.gsmworld.com/news/press_2007/press07_22.shtml The whitepaper is available at: http://www.gsmworld.com/documents/nfc_services_0207.pdf

18 Juels, A. RFID Security and Privacy: A Research Survey. *IEEE Journal of Selected Areas in Communications*, 24 (2), 381-394, 2006.

19 *The Use of RFID for Human Identity Verification*. Department of Homeland Security, Data Privacy & Integrity Advisory Committee, Adopted December 6, 2006. (Report No.2006-02) Available at the DHS homepage: http://www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_12-2006_rpt_RFID.pdf

20 *The use of RFID for Human Identification*. A draft report from DHS Emerging Applications and Technology subcommittee to the Full Data Privacy and Integrity Advisory Committee, Version 1.0. Department of Homeland Security, 2006

This document is a subcommittee input to [DHS]. Available at the DHS homepage: http://www.dhs.gov/xlibrary/assets/privacy/privacy_advcom_rpt_rfid_draft.pdf

21 The Register. *US.gov tunes out scathing RFID privacy report*. Published at http://www.theregister.co.uk/2006/11/02/rfid_study_disavowed/

22 The Register. *California Senate fights RFID tracking for schoolkids*. Published at http://www.theregister.co.uk/2007/04/17/california_fights_rfid_child_monitoring/

For a presentation of Geir M. Køien, please turn to page 3.

Privacy Preserving Data Mining in Telecommunication Services

OLE-CHRISTOFFER GRANMO, VLADIMIR A. OLESHCHUK

Ole-Christoffer Granmo is Associate Professor at Agder University College, Norway

Vladimir A. Oleshchuk is Professor of Computer Science at Agder University College, Norway

Privacy preserving data mining is a new research direction in data mining and knowledge discovery. A main reason for the rapid development of this research area is the growing awareness of the accumulation of huge amounts of easily available data on the Internet – data that may involve a threat to the privacy of users. The problem of user privacy is even more critical for telecommunication applications since they enable user mobility, which in turn may involve accumulation of sensitive spatial data. Such data can for instance reveal driving habits, etc. In this paper we give an overview of selected approaches in the area of privacy preserving data mining, with special focus on approaches that are suitable for developing privacy preserving applications in the area of telecommunication.

Introduction

Privacy preserving data mining is a new research direction in data mining and knowledge discovery [VC04]. A main reason for the rapid development of this research area is the growing awareness of the accumulation of huge amounts of easily available data on the Internet – data that may involve a threat to the privacy of users. Privacy can for instance be threatened when data mining techniques are used to connect personal identifiers such as addresses, names, etc., with other, more sensitive, person related information. As an example, shopping or travel habits could be mined from the traces of information that remain after Internet use, thus potentially revealing sensitive information. On the other hand, if privacy concerns can be addressed, society may also benefit from the knowledge that can be distilled from sensitive information. Different approaches have been developed to tackle this dilemma. Some approaches use data obfuscation that modifies original personal data in order to protect a person's privacy. Another direction is to develop new privacy preserving data mining algorithms that protect a person's privacy in the sense that no private data is revealed, for instance by using secure multi-party computations [Pin02].

In this paper we give an overview of selected approaches in the area of privacy preserving data mining with special focus on approaches that have potential for application in the area of telecommunication.

Data Mining

Data mining concerns the extraction of knowledge from potentially large collections of unstructured and structured data, such as medical records, telecommunication customers' calling data, and web discussion forums. Basically, extracted knowledge consists of discovered patterns and correlations that are hidden in the data. In this sense, data mining can be said to add *new meaning* to data. Searching patient data from all hospitals in the world for patterns could for instance uncover new relations between potential treatments and outcomes, symptoms and diseases, and so on.

A number of distinct approaches to data mining have been identified [DHS01, Mi97], namely, classification, association rule learning, clustering, and multidimensional scaling (cf. data visualization). These different techniques can be summarized as follows:

- *Classification techniques* support categorization of elements within a data set into predefined categories. A classifier could for instance classify e-mail as being either from the category 'desired' or from the category 'spam' [SDHH98].

- *Association rule learning* concerns the discovery of co-occuring data elements (cf. events), including uncovering of causal relationships. In telecommunications, association rule learning can for example be used on calling detail data to identify pairs of customers that frequently call each other (which in turn can be used to identify so-called calling circles) [We05].

- *Clustering techniques* partition a data set into subsets so that the elements within each subset are similar; that is, they share common traits. Customers could for instance be partitioned into subsets such as 'Business Customer' and 'Residential Customer', based on their calling behavior. Identifying customer groups with similar behavioral traits may in turn enhance or enable directed marketing [We05].

- *Multidimensional scaling techniques* are used to detect meaningful underlying dimensions in a high-dimensional data set. Such techniques are often used in data visualization to uncover similarities among elements within the data set at hand. For example, so-called scatter plots can be used to

visualize network traffic for intrusion detection purposes [Ma01].

In the telecommunications industry, as discussed in [We05], huge amounts of data are produced and stored, including:

- Call data describing the calls that traverse the telecommunication networks,
- Network data concerning the state of hardware- and software components,
- Customer related data.

Within such tremendous amounts of business critical data, valuable knowledge may be hidden. Indeed, as seen above, it has previously been demonstrated how data mining can be used to discover new patterns and correlations within such sets of data. Applications include telecommunication fraud detection, improving market efficiency, and fault detection and localization.

Privacy Concerns and Data Mining

In practice, the applicability of data mining is limited by *privacy concerns*. The nature of data mining is to uncover hidden patterns and correlations that are not explicitly given. As such, data mining also has potential for uncovering sensitive information that concerned parties consider private.

As discussed in [We05], when the telecommunication company MCI launched its 'friends and family' campaign[1)] in 1991, customers had to report their calling circles themselves in order to benefit from the campaign. It would arguably have been more effective to proactively offer calling circle deals by identifying calling circles automatically by mining call detail data. However, uncovering such calling circles automatically, and using them for marketing purposes, could anger customers. After all, calling circles can contain private and sensitive information.

Another example is using data mining techniques to detect disease outbreaks, which may require data on disease incidents, patient background, and so on [XCL06]. Since such data is sensitive, legal concerns may hinder their free use, even for beneficial purposes. In other situations, the target data may be partitioned across several organizations, thus organizational policy- and commercial concerns may restrict data use, in addition to legal ones. Apparently, in both of the above examples, some kind of *privacy preserving* data mining is needed.

Privacy-Preserving Data Mining

An important research question in data mining is: *to what degree can we have both data privacy and the benefits of data mining?* [Mi06]. From a first glance, choosing between privacy and the benefits of data mining seems unavoidable. However, in recent research, the two have in certain cases been combined. The challenge lies in getting valid data mining results without learning the underlying data values. As we shall see in the following, previous statistical work on data disclosure and recent cryptographic techniques form the basis for current solutions.

Data Model

When studying privacy-preserving data mining it is useful to consider how data may be partitioned among the involved parties. In some cases, organizations may collect the same kind of data about different entities (for example people, traffic, etc.). From a database perspective, we may then say that the data is partitioned *horizontally*; that is, the same schema is used to store the data at each site. In other cases, organizations may organize data using different schemas, meaning that they collect different kinds of data, perhaps on the same entities. We then say that the data is partitioned *vertically*.

Figure 1 shows one example of horizontal and vertical partitioning of data among two telecommunication companies and a third party. Consider the situation where the two telecommunication companies plan a joint system for detecting misuse of mobile phones (fraud detection). For such purposes, the two companies each maintain a database with daily aggregated customer calling data. Database 1 and Database

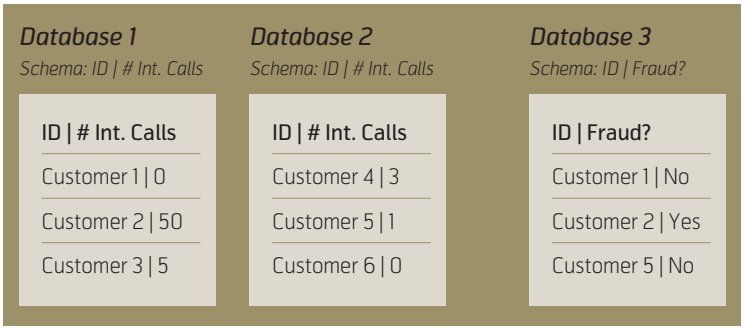


Figure 1 Database 1 and Database 2 partition a data set horizontally (same schema, different entities). Database 3 introduces vertical partitioning (different schema, same entities)

1) *Reduced rates within small calling circles consisting for example of friends and family members.*

2 in Figure 1 contain one such aggregated quantity, namely the number of international calls associated with each customer for a given day. Finally, assume that information on previous frauds is stored in another database (database 3 in the figure), maintained by the third party.

The complicating factor in this scenario is the following: *None of the companies want to reveal their customer data, and legal restrictions do not allow joining of Database 3 with Database 1 and Database 2. Yet, it is desirable to construct rules for detecting frauds using the joint data*. Note that we will use this example throughout the paper to illustrate various privacy preserving data mining concepts.

Main Techniques

There are essentially two main approaches to achieving privacy preserving data mining. We will now briefly give an overview of these before we discuss specific privacy preserving techniques for classification, association rule learning, clustering, and multidimensional scaling. The two approaches can be summarized as follows:

- *Data transformation/randomization* essentially amounts to modifying sensitive data so that it loses its sensitive meaning, but still retains statistical properties of interest. For example, one might be interested in mining the aggregated, statistically significant properties a collection of data elements possesses, while the owners of the data elements need to protect their data. Privacy can then be preserved by only revealing randomized and transformed versions of the data, for instance perturbed using a randomization algorithm that maintains statistical properties.

- *Secure multi-party computation* is a computation performed by multiple parties where each party has in its possession a part of the input needed to perform the computation. However, at the end of the computation, the parties should only have learned the result of the computation (in addition to the input the party itself provided to the computation). Secure-multi-party computation methods used for privacy-preserving data mining are typically based on such building blocks as secure sum, secure set union, secure size of set union etc. [Pin02]. For example, secure sum protocol describes how to calculate the sum of distributed items without revealing their true values. More details can be found in [CKVLZ]. (See also the paper on secure multiparty computations in this issue, page 20-26.)

The above main approaches form the basis for more specific data mining techniques, as overviewed in the following.

Privacy Preserving Classification Techniques

Constructing a classifier typically involves so-called *training*. In brief, training means using already categorized data elements, called *training data*, to derive a procedure for classifying new and previously unseen data elements. For instance, one might want to learn a procedure for predicting "Fraud"-attempts based on the "Number of International Calls" of customers, using the data in Database 1, Database 2, and Database 3 in Figure 1 as training data. One approach could be to estimate the parameters of a parametric model, so that the model predicts as accurately as possible (for example using a maximum likelihood based approach).

Privacy preserving classification techniques essentially address two questions:

1 How can we train a classifier without revealing the training data itself?

2 How can we classify new unseen data elements without revealing those data elements?

It turns out that in most cases each type of classification technique requires a tailored solution, targeting either vertically or horizontally partitioned data, and using data transformation/randomization and/or secure multi-party computations. For instance, in Privacy Preserving Nearest Neighbor Search for horizontally partitioned data [XCL06], the goal is to find the training data element that is nearest to the data element that we want to classify; however, without revealing any of the data elements, only the final classification.

Similarly, a Naïve Bayesian classifier for vertically partitioned databases may need to be trained in settings where only some of the parties know the class attributes of the training data [VC04]. Thus, in the worst case, only one party knows the class attributes of the training data while the other parties know the other attributes (cf. the fraud detection example from Figure 1). In brief, the problem is that training requires both the attributes and the class of data, yet privacy concerns may not allow this information to be shared.

Privacy Preserving Association Rule Learning

The goal of association rule learning is to find specific patterns that represent knowledge in generalized form without referring to particular data items.

Because of this one might say that association rule learning only represents an indirect threat to privacy. However traditional methods require access to the data set in order to be able to find association rules.

In the case presented in Figure 1 we assume that the databases are owned by different parties and no one wants to disclose their data to other parties. Again, the main concern is how to avoid revealing data to other parties. Both data transformation/randomization and secure multi-party computation techniques have been applied to develop privacy preserving methods for association rule learning.

In the first approach [ESAG04] data are randomized such that the true values cannot be estimated with sufficient precision. The typical problem can be formulated as following. Assume that there are several customers having databases containing private information and one server which is interested in learning association rules based on statistically significant properties of this distributed information. Customers protect the privacy of their data by perturbing data with some randomization algorithm and then submitting the randomized version of data to the server. The framework for mining association rules from data that have been randomized to preserve privacy is described in [ESAG04].

The second approach assumes that data are distributed between two or more sites and the purpose is to learn global association rules without revealing data to other sites. This approach was applied to both horizontally and vertically partitioned data [KC04, VC02].

Privacy Preserving Clustering Techniques

The goal in clustering is to partition data elements into clusters so that the similarity among elements belonging to the same clusters is high, and so that the similarity among elements from different clusters is low. In privacy preserving clustering a main goal is to find the clusters in the data without revealing the content of the data elements themselves. For example, two or more companies may decide that performing clustering on customer data might improve their directed marketing efforts. However, they are not willing to reveal their own customer data to the other party. Again, note that the data may be partitioned vertically and/or horizontally among the involved parties. To exemplify, in [CKVLZ] a scheme for privacy preserving of so-called EM-clustering is proposed that only reveals aggregated quantities, using multi-party secure computations. In [OZ03], data transformation/aggregation is used, so that the clustering performed on the distorted data is still valid.

Multidimensional Scaling Techniques

Informally, Multidimensional Scaling (MDS) is the process of transforming a set of points in a high dimensional space to a lower dimensional one while preserving the relative distances between pairs of points. This property is important in the context of data visualization where it is important to preserve relative relationship between data items while reducing dimensionality. The privacy preserving visualization problem for this case can be formulated as follows. Assume that Alice and Bob have two sets of private data items that represent sets of points in *m*-dimensional Euclidian space. They wish to visualize jointly both sets without revealing data items in such a way that none of the visualized points can be correlated with items in the original data sets and therefore it is not possible to find the origin of any point.

Existing Problems and Future Research

As indicated in the above overview, the area of privacy preserving data mining encompasses a number of novel and intriguing problems that span several fields. For the field of telecommunications, we consider the following recent and more or less unexplored avenues of research to be particularly important:

- *Privacy preserving adaptive control* concerns how to learn to control a dynamic system without revealing your intentions, and without revealing the internal system properties. One classical example of adaptive control is multiplayer multi-armed bandit problems with unknown reward distributions [NT87].

- *Privacy preserving adaptive resource allocation* concerns the problem of allocating resources to entities in order to optimize some performance criterion, without revealing information about the resources or the involved entities.

- *Privacy preserving adaptive routing* concerns how routers can decide upon near optimal routing strategies, without revealing information on customer preferences, traffic patterns, etc.

- *Privacy preserving multi-dimensional scaling:* One of the possible ways to address privacy preserving multi-dimensional scaling would be an approach based on the Johnson-Lindenstrauss lemma [JL84] which notes that any set of s points in m-dimensional Euclidean space can be embedded into k-dimensional subspace, where k is logarithmic in s, such that the pair-wise distance of any two points is maintained within an arbitrarily small factor.

Therefore, by projecting the data onto a random subspace, we can dramatically change its original form while preserving much of its underlying distance-related statistical characteristics.

Another problem of interest is the application of privacy preserving techniques when mining data in video information received from distributed monitoring cameras or in measurement data received from distributed sensor networks. In this context privacy-preserving pattern matching is an important problem [Ol]. We are presently pursuing research problems from several of the above problem categories, with a particular emphasis on so-called Learning Automata based routing and resource allocation [OMG07, GOMO07] as a foundation for privacy preserving decision making, and on distributed Bayesian techniques [EGEM06] as a foundation for privacy preserving video analysis.

References

[CKVLZ] Clifton, C, Kantarcioglu, M, Vaidya, J, Lin, X, Zhu, M Y. Tools for privacy preserving distributed data mining. *SIGKDD Explor. Newsl.*, 4 (2), 28-34, 2002.

[DHS01] Duda, R O, Hart, P E, Stork, D G. *Pattern Classification*. Wiley-Interscience, 2001.

[GOMO07] Granmo, O-C, Oommen, B J, Myrer, S-A, Olsen, M G. Learning Automata-based Solutions to the Nonlinear Fractional Knapsack Problem with Applications to Optimal Resource Allocation. In: *IEEE Transactions on Systems, Man and Cybernetics, Part B*, 37 (1), 166-175, 2007.

[JL84] Johnson, W B, Lindenstrauss, J. Extensions of Lipshitz Mapping into Hilbert Space. *Contemporary Math.*, 26, 189-206, 1984.

[ESAG04] Evfimievski, A, Srikant, R, Agrawal, R, Gehrke, J. Privacy preserving mining of association rules. *Information Systems*, 29, 343-364, 2004.

[KC04] Kantarcioglu, M, Clifton, C. Privacy-Preserving Distributed Mining of Association Rules on Horizontally Partitioned Data. *IEEE Trans. Knowledge Data Eng.*, 16 (9), 1026–1037, 2004.

[Ma01] Marchette, D J. *Computer Intrusion Detection and Network Monitoring*. Springer, 2001.

[Mi97] Mitchell, T M. *Machine Learning*. McGraw-Hill, 1997.

[NT89] Narendra, K, Thathachar, M. *Learning Automata: an introduction*. Prentice-Hall, 1989.

[Ol] Oleshchuk, V. Privacy preserving pattern matching on sequences of events. *International Journal of Computing*, 4 (3), 85-90, 2005.

[OZ03] Oliveira, S, Zaiane, O R. Privacy Preserving Clustering By Data Transformation. In: *Proceedings of the 18th Brazilian Symposium on Databases (SBBD 2003)*, Manaus, Brazil, 6-8 October 2003, 304-318.

[OMG07] Oommen, B J, Misra, S, Granmo, O-C. Routing Bandwidth Guaranteed Paths in MPLS Traffic Engineering: A Multiple Race Track Learning Approach. To Appear in *IEEE Transactions on Computers*. (Accepted January 23, 2007)

[Pin02] Pinkas, B. Cryptographic techniques for privacy-preserving data mining. *SIGKDD Explor. Newsl.*, 4 (2), 12-19, 2002.

[SDHH98] Sahami, M, Dumais, S, Heckerman, D, Horvitz, E. A Bayesian Approach to Filtering Junk E-Mail. In: *Learning for Text Categorization: Papers from the 1998 Workshop*. AAAI Technical Report WS-98-05, 1998.

[VC02] Vaidya, J, Clifton, C. Privacy preserving association rule mining in vertically partitioned data. In: *Proceedings of the Eighth ACM SIGKDD international Conference on Knowledge Discovery and Data Mining*, Edmonton, Alberta, Canada, 23-26 July 2002. KDD ‘02. New York, NY, ACM Press, 639-644.

[VC04] Vaidya, J, Clifton, C. Privacy-Preserving Data Mining: Why, How, and When. *IEEE Security and Privacy*, 2, (6), 19-27, 2004.

[VC04] Vaidya, J, Clifton, C. Privacy Preserving Naive Bayes Classifier for Vertically Partitioned Data. In: *Proceedings of the 2004 SIAM International Conference on Data Mining*, Orlando, 2004.

[We05] Weiss, G M. Data Mining in Telecommunications. In: *Data Mining and Knowledge Discovery Handbook: A Complete Guide for Practitioners and Researchers*. Kluwer Academic Publishers, 2005, 1189-1201.

[EGEM06] Wold Eide, V S, Granmo, O-C, Eliassen, F, Michaelsen, J A. Real-time Video Content Analysis: QoS-Aware Application Composition and Paral-

lel Processing. In: *ACM Transactions on Multimedia Computing, Communications, and Applications (ACM TOMCCAP)*, 2 (2), 149-172, 2006.

[XCL06] Xiong, L, Chitti, S, Liu, L. Nearest Neighbor Classification across Multiple Private Databases. In: *Proceedings of CIKM'06*, ACM, 840-841, 2006.

Ole-Christoffer Granmo obtained his MSc in 1999 and PhD degree in 2004, both from the University of Oslo, Norway. He is currently Associate Professor in the Department of ICT, Agder University College, Norway. His research interests include Intelligent Systems, Stochastic Modelling and Inference, Machine Learning, Pattern Recognition, Learning Automata, Distributed Computing, and Surveillance and Monitoring.

email: ole.granmo@hia.no

For a presentation of Vladimir A. Oleshchuk, please turn to page 3.

Distributed Health Records, Cryptographic Pseudonyms, and Privacy

STIG F. MJØLSNES

Stig F. Mjølsnes is Professor at Department of Telematics, Norwegian University of Science and Technology

The public health administrations collect records of patients systematically and extensively, thereby constructing databases that will enable medical researchers easy access to information vital for conquering diseases, and enable managers to run health services optimally. On the other hand, the realization of nationwide databases of this type, and keyed to person identity, may be prohibited by legal and privacy considerations. Addressing the technicalities and utility of this long standing question, a secure multiparty system structure that alleviates this ambivalence is presented here.

1 Introduction

1.1 Motivation

Health administrations and epidemiological researchers systematically accumulate personal health information in regional and national computer based registries. Acquisition, storage, inspection and usage, maintenance and management of this kind of data must be governed by national and international rules and regulations. Privacy of health data is considered a major user requirement, and supportive technical means should be constructed.

1.2 Part of Norwegian History

The motivation and many of the results reported in this paper originated with work[1)] carried out in the process of technology advice to a Norwegian governmental committee supported by the Norwegian Ministry of Health (Boe-committee[2)]) mandated to report on and propose new laws and regulations for national health registers. My concrete technological solutions, also to be presented here actually, premised reasonable and constructive requirements to the more general policy discussion in the committee and beyond, with respect to national law and regulation advancing the protection of personal health information registrations [1]. A long public, often political discussion followed with changing governments and sectorial interests. Many of the concepts are now part of Norwegian Health Registry Act of May 2001 and the regulations. Some essential concepts developed by the lawmakers are *personidentifiable registry* (indexed by public birthnumber and name); *person-unique fields* (fields that enable (time independent) linking of records to the same person); *encrypted personidentifiable registry* (person-unique fields encrypted); *de-identification* (deleting person-unique fields). However, technically satisfactory systems and best practice of actual pseudonymous health registries have not been established in Norway yet.

1.3 Architecture Overview

This paper presents an architecture based on multiple autonomous database management systems, or local registers (*LR*), each set up in the vicinity, but organizationally separate from a regional hospital. The hospital records the patient's health attributes by his or her identity number, normally a unique person number issued at birth. A selection of the recorded attributes is forwarded to *LR*, but now keyed to a cryptographic pseudonym mapped from the identity number. The pseudonymization computation in the registration process must be performed in cooperation with the patient. The patient provides the pseudonym and a proof (a mathematical witness) of his pseudonym (previously issued to the patient by a trusted center) to a tamperresistant verification unit at the hospital. This trusted hardware unit is able to verify the correspondence between the identity (provided by the hospital) and the pseudonym (provided by the patient) without revealing any more information from the computation. The pseudonym, encrypted under *LR*'s public key, is revealed to the hospital and sent to *LR*. As a result, the hospital will not know the patient's pseudonym, and *LR* will not know the patient's identity. The hospital cannot use the pseudonym verification unit as an efficient pseudonym oracle because no computation will be performed unless both the identity number and the corresponding pseudonym are input. Each individual is enabled to inspect their records anonymously by presenting the pseudonym and prove the knowledge of the witness. Furthermore, a re-identification (from pseudonym to name) is made possible within the system architecture, but conditioned on a message response given by the person concerned. Hence, the identity of the person is never revealed to *LR*, nor is it necessary

1) *This work originally started in 1990.*

2) *The Governmental committee 1990-93 on privacy and health records established by the Norwegian Ministry of Health autumn 1989 with professor Dr.Jure. Erik Boe as chair.*

 ISSN 0085-7130

to reveal the pseudonym to the hospital. The distributed system of databases provides extensive but controlled online access to a variety of health statistics without releasing individual attributes. Database queries are answered with locally aggregated data prepared by query servers, and subsequently collected and combined by the requesting workstation. Statistical inference control methods may be employed in addition to standard access control methods to execute the security policy at the query servers. Means are also provided, on a continuous basis, to join medical research records (identity-keyed) with registered health records (pseudonymkeyed), conditioned on active acknowledgment by the patient, without giving away the pseudonym to the researcher nor revealing the identity number to the database administration.

1.4 Related Literature

The technical challenges dealt with in this paper stem from political and organizational issues. On the political level, the reporting of the Boe-committee on privacy and national health registers is documented in [2]. A quite lively public discussion ensued that report, and in fact continues at the time of writing this. A public report [3] issued by the Norwegian Ministry of Health appeared in December 1998 and summarizes the process of additional committee reports and the long hearing period for the law proposed. Additional reports were issued in 2004 and 2005 from the Norwegian Ministry of Health.

Recently, by wide scope and popular approach, Garfinkel's book [4] describes a variety of privacy problems raised by the existence and use of medical databases, perhaps mostly seen from a US perspective.

Somewhat dated, Denning's book [5] is still relevant as an introduction to database security issues. More recent database security research is reported in [6], but that direction of work is primarily motivated by military requirements, while best practice for information technology deployment and operation in health institutions currently emphasizes the importance of structural flexibility rather than strict operational hierarchy.

The state of the art in distributed database systems can be found for example in the recent edition of the textbook of Özsu and Valduriez [7]. However, the weak understanding that we currently have of how to achieve *multiparty* security in distributed database systems is of course reflected in this and other contemporary textbooks. The term multiparty security designates that security requirements to the system and its service are at variance with respect to distinct user categories or roles accepted by the system. The system discussed in this paper presents multiparty security requirements and solutions. In the proposed system, the core cryptographic mechanisms for generating pseudonyms are based on techniques established in the paper by Brandt et al. [8].

A brief report at Asiacrypt [9] presented an outline of the ideas of the proposals in this paper. A more substantial technical report, written in Norwegian, was included in [2].

1.5 Outline of Paper

The rest of the paper is organized as follows: Section 2 defines the essential problems and assumptions. Section 3 constructs the conceptual model, the system architecture and detailed functionality. Section 4 describes the protocols for the pseudonym computations. Section 5 details how the patient is in control of the registration and verification of data records. Section 6 proposes how to manage an operational transition from legacy registries to new registries employing pseudonym indices. Section 7 discusses security and the statistical inference problem. Section 8 carries out an analysis of various user modes of the proposed system. Section 9 discusses implementation feasibility, and Section 10 concludes the paper.

2 Problem Description

A simplified model of the problem is the following. A hospital's registry records the patient's medical information. Each record is indexed by a unique identifier key, which typically includes the national identity number or social security number of the patient. In general, central national registries and local registries for health administration and research already exist. The main purpose of these registries is to contain and provide personal medical information for statistical use, for instance in epidemiological investigations.

On the one hand, indices that uniquely identify the individual patients are necessary for all such registries, not only for the purpose of medical treatment, but also because it makes it possible to maintain records at the individual level. Information from several registries can easily be merged on an individual level resulting in a firm database for statistical and other types of investigations.

On the other hand, reasonable privacy rules and regulations will require that such medical databases and registries operations will not release information that can be unconditionally linked to a given person.

Requirements to the standard categories of information security, including confidentiality, authenticity, integrity and availability, can be identified in the sys-

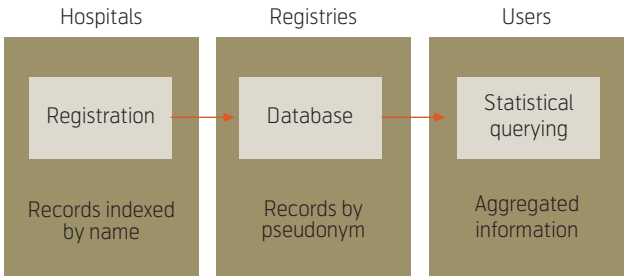


Figure 1 The conceptual model with three distinct organizational domains. The information flow is uni-directional from registration to query domain

tem design presented here. Security measures covering these requirements will be:

- Access control mechanisms to the computers and the data stored therein;
- Protection against modifications of software and data in the computers;
- Confidentiality and integrity of the computer communications;
- Digital signatures for non-repudiation, receipts and auditing.

Security mechanisms that satisfy such requirements are mentioned many times throughout the description of the system, but are simply assumed available without further detailed constructions. The construction and application of these mechanisms are very important for the totality, but this paper focuses on the security mechanisms particularly required for providing privacy of medical health registries.

A statistical database should be able to offer as many types of statistical queries as possible with respect to any *group* of individuals. This implies that health information must be recorded individually. At the same time, the aim is to protect the individual health information as much as possible. This implies that not all queries to the database can be responded to, but must be differentiated with respect to the user's authorization. Moreover, a clever combination of statistical queries may reveal individual health data, and as such must be controlled.

3 System Architecture

3.1 Model and Security Policy

The general model of the proposed system is a three tiered structure as shown in Figure 1. The data acquisition and registration are represented by the block on the left. This input will comprise selected health data from the hospital's patient records, either new entries or corrections of already entered data.

The centre block represents the database and its management. In the detailed design model, this block will consist of several registries associated with local registry organizations.

The block to the right represents the statistical output process, where authorized users make statistical queries to the health database. The users may be geographically dispersed, perform access independently of the location of the registry, and they may have access only to parts or to the whole database.

The solution strategy is to reduce and remove the *direct* informational links between the individual's identity string and the corresponding health data attributes as the data 'flow' from the registration source to the user. The hospital will register and store the health information by personidentifiable data, such as the patient's full name and social security number. However, the registry organization must use cryptographically generated pseudonyms. This means that the registry organization will possess the individual records, but not be able to link these records to the identity of any one person.

Furthermore, the registry organization can choose to respond to queries in the form of aggregated data only, and not release data from any individual record. For instance, replies to statistical queries could be cast in such a way that it will not reveal health attributes of any individual patient, while still satisfy the statistical requirements expected by the user.

Other users could be authorized to read excerpts of individual health information within a special group of patients. The registry will release the individual health information to these users, for instance indexed by some pseudonym ordering.

Some registry users could require to follow up individual patients with respect to recorded health data. Even though the records are indexed by cryptographic pseudonyms, it is still technically feasible within the proposed system to efficiently link from pseudonym to identity, by allowing the operation of a trusted third party.

3.2 Detailed Model

3.2.1 Structure

Refer to Figure 2 in the following. A local general health registry associated with a hospital is denoted *LR*, and a special registry (for instance cancer reg-

istry, medical birth registry) is denoted *SR*. A query processor connected to an *LR* or an *SR* is denoted *LQ*. Without further details, the assumption is that the source of identitybased health information is the hospital's patient record system of some kind, denoted *JR*. For the purpose here, it is reasonable to assume that all *LR* are structured according to a standardized database scheme, but the data structure of the special registries *SR* will be different.

A uniquely identifying number or code institutionally linked to an individual will be denoted *Id*, such as a personal number or social security number. The pseudonym *Pid* is a number or code that is cryptographically linked to an *Id*, as will be shown later. Let *A* denote health data and descriptions that can be attributed to any individual patient. Let an *LR* store a table where each record contains an identity and the corresponding health attributes, denoted (Pid, A_1). Hence, the data attributes of the health records are kept to an absolute minimum, in order to abstract away from details not directly relevant to the problems addressed here. Let *SR* store a table where each row contains an identity and the corresponding health attributes, denoted (Id, A_2). Hence, both types of registry relate attribute values to distinguished indices uniquely identifying the patient's person-unique fields. However, to join attributes from *LR* and *SR* on identity, we must first be able to compute the relation (*Id*, *Pid*).

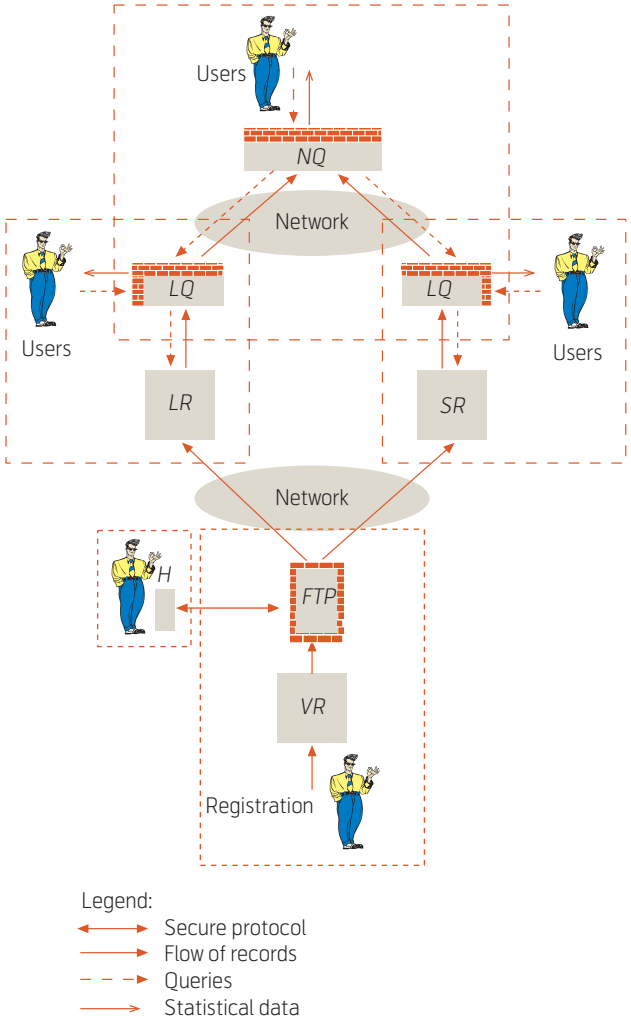


Figure 2 System architecture that shows one local registry instance LR and one special registry instance SR. The dotted rectangles depict organizational domains

3.2.2 Input

The hospital's *JR* stores the patient health data tables with person-unique fields as part of the primary key. The *JR* is assumed to be within the domain of the hospital organization and its corresponding security policy with respect to access and information flow control. This policy will also determine which data in *JR* that will be reported to *LR*, including the procedure for how this is carried out.

The health records to be input to *LR* or *SR* will be pseudonymized with the help of a *tamper-resistant pseudonym tester TPT*. This *TPT* unit must be located within the hospital organization because it is used in the registration of the patient's health data. Hence, this computing unit is required to be tamper-resistant so that *nobody* is able to modify the functionality or read out temporarily data stored inside. Consequently, the *TPT* unit can be securely located within the hospital territory without misuse. Computing pseudonyms for given identities can only take place if *JR* and *LR* cooperate in computing a pseudonym, as described in Section 4.

The communication unit that contains *TPT* receives messages from *JR*, processes these messages with the help of *TPT*, then forwards the messages. The correct recipient is either a local register *LR*, or a special register *SR*, or perhaps both. The recipient address is determined by *JR* and attached to the message.

Normally, data communication protocols provide the sender address to the receiver. In principle, a system administrator of *SR* can easily determine which hospital the patient report is coming from. So if it is required that the identity of the reporting hospital should be hidden, then the sender address cannot be sent to *SR*. One possible solution is to introduce a trusted intermediate communication node *MIX* that removes the original sender address. The message authentication will take place in two steps, first between the *TPT* and the *MIX*, then between the *MIX* and the *SR*. It should be sufficient with a single *MIX* in the system.

A theoretically more elegant mechanism for solving this problem is to use *group signatures* [10]. All *TPT* will be enabled to digitally sign the message according to this cryptographic protocol. Any *SR* can verify that the message originated from the set of authorized *TPT* without revealing the exact identity. If, at a later stage, it becomes necessary to know which *TPT* was the sender, the signature can be 'opened' by all *TPT* taking part in an opening protocol.

The main task of the *TPT* is to *verify* the computational correspondence between the identity code, received from *JR*, and the pseudonym code received from the patient. If this turns out positive the identity index is replaced with the pseudonym index and the message is forwarded. The pseudonym code is the computational result of a cryptographic transformation of the identity code, precomputed as described in detail in Section 4. Importantly, the *TPT* is only able to compute the required verification procedure if supported by the patient's *H*, a very small computing token that can store, among other data, the witness number of the patient's pseudonym. This process of pseudonymization is described in detail in Section 4.3.

3.2.3 Output

The users of the health data extract information from *LR* employing the *LQ*. The *LQ* provides a statistical query service based on the database *LR*. The users are granted access rights (authorized) according to a specified security policy, ideally by which the *LQ* machine will control information access, flow and statistical inference.

A central nationwide query processor *NQ* will perform a service similar to *LQ*. The difference is that *NQ* service can be based on all local registries *LR* and *SR* combined. Each user of *NQ* is granted access rights according to a well defined security policy, by which the *NQ* machine will control information access, flow and statistical inference, analogous to the functionality of *LQ*.

Note that *NQ does not* require raw records if the data structures of the registers are the same or similar. This is discussed further in Section 8.5.

If the registry data structures (schemas) are not homogeneous, integration into a cohesive distributed database will pose a challenge. In fact, a number of the *SR* are legacy systems and developed independently of each other. Accordingly, there will exist types of statistical queries that *NQ* cannot answer solely based on statistical data from the *LQ* processors, but will require the 'raw data' (the records indexed by pseudonyms) combined from different *LR* and *SR*. This means that raw data must be sent to *NQ* if such queries can be answered. This does not imply that *NQ* must maintain a central database, but that *NQ* must *temporarily* collect the raw records from the distinct *LR* in order to be able to compute the query. Subsequently, the raw records submitted must be deleted. Further optimization to this process seems to be possible, but is not pursued within this paper.

3.2.4 Security Management Domains

Clearly, it is reasonable to put the pairs of *LQ* and *LR* within the same organizational domain. And similar with *LQ* and *SR*. Users restricted to one registry can then be managed by this local domain and its security policy. Another advantage is that this boundary aligns perfectly with already established and running *SR* organizations.

The *NQ* services demand its own organizational domain that can manage user authorization and debiting, establish and maintain links to the local and special registries, and coordinate the standardization of database structures and data communications. Other tasks that naturally fall within the authority of a central *NQ* are coordination of a harmonized security policy and quality control of data registration.

Note that technically speaking, a centralized management of the local machines is completely feasible and will probably turn out to be very cost-effective. More precisely, it is technically feasible to employ centralized system management sustaining a local authorization and security policy.

4 Pseudonymization

4.1 Security Assumptions and Objectives

Our primary concern is that *LR does not* get knowledge about the relation (*Id*, *A*). What are the assumptions we have to rely on to satisfy this requirement? First we examine the assumptions necessary for the organizations of *JR* and *LR* with respect to the pseudonymization process.

Certainly, we assume that *JR does not* reveal information about (*Id*, *A*) to outsiders. This brings the assumption that *LR* does not have any access to *JR*. It is reasonable to assume that the pseudonyms do not need to be kept secret with respect to *JR*, because the register *JR* already has access to all health information of the patient. Moreover, we make the assumption that *LR* will not leak any information to outsiders. This implies that the problem can be reduced to the problem of how *JR* can send information on the format (*Pid*, *A*) to *LR* or *SR*.

The foregoing assumptions are probably too strong in practice because there will likely be individuals associated with both the *JR* and the *LR,* for instance a medical doctor carrying out research projects. Simple access to pseudonyms at *JR* results in easy re-identification that cannot be controlled seen from the *LR* side. This argument implies that we must strengthen the security requirements of the pseudonymization process such that neither *JR* nor *LR* will know or be able to compute the informational link (*Id*, *Pid*).

It follows that the process of pseudonymization must be performed by others than the *JR*-organization. The immediate solution is to let the patient himself, or someone on his behalf, compute the link between identity and pseudonym. Therefore the clarified objective becomes: each patient (or his or her representative) shall be enabled to compute one and only one pseudonym linked to the identity.

The requirement that neither *JR* nor *LR* shall be able to link identity and pseudonym must be supported by the assumption that the registries *JR* and *LR* or *SR will not cooperate* in linking the identity with pseudonym. It is straightforward to link these by cooperative computing, for example by *JR* somehow coding the *Id* into the health records that are sent to *LR*. Information about *Id* can for instance be 'coded' into an attribute variable without any possibilities for third parties to check this. (Basically, this is the problem of subliminal or side channels in cryptographic literature.)

Furthermore, the patient must have some method of inspecting *LR* to verify that the information stored is timely, relevant, correct and complete. As already stated, the registry *LR* shall not be able by itself to link identity and pseudonym; this individual inspection control must be performed *anonymously*. Moreover, it is important that each patient only has access to *LR* records concerning himself. Hence, some mechanism for authentication of the claimed pseudonym must be provided for.

4.2 Patient Computer

Either the patient, or his or her representative, must be equipped with some computing 'help' in order to be able to participate in the computation of the pseudonym, and other protocols. This computing help could be from a small computer device, such as a smart card, or it could be part of the functionality of an electronic wallet or wristwatch, mobile phone, or other wearable computer device. Formfactor, size, price and features may vary, but the computing device must be able to carry out public key computations efficiently. It must also be able to communicate with the hospital's *TPT* unit, most preferably wireless. The patient allows the device to engage in communication with the *TPT*, the rest will be performed automatically by the machines. Some indication of the processing and finishing should be shown by the device, as assuring feedback to the patient. The detailed construction of the computing device is outside the scope here, so in the following this computing device will simply be denoted *H*.

There are several important characteristics for the management of *H*. It will *not* be issued and owned by a particular organization, and will not contain any information secret to the holder, along the same principles that you would expect for your own wallet. At this point, note that this approach is markedly different to most of the contemporary solutions proposed in smartcard schemes of banks, identity cards and so on, where it is an organization or a service provider or an institution that 'lends out' the user instrument to the client, and where the security is based on the secret information stored in the tamper-resistant device. The *H* device is the patient's 'information wallet' that can be used in the communication with other organizations and services, as needed. An organization simply needs to issue the *information* that the client needs in order to start using the services offered. This solution secures *both* the client and the organization interests, whereas most of the current 'instrument' solutions are one-sided in that they only take the organization's security into view. Suitable devices that can realize *H* are commercially available today in the form of mobile phone, pocket computers and supersmartcards.

If the owner is afraid that *H* might be lost, stolen or misplaced, then the access to the information can be protected by some password, or even better, by biometrics such as thumbprint or speaker dependent voice recognition. The data stored in *H* can be backed up, either employing a home PC or with some trusted third party service.

4.3 A Cryptographic Transformation

A basic cryptographic transformation that can be applied in the system is presented in Ref [8], and is recaptured here for completeness. The following requirements must be satisfied for the transformation *T* from *Id* to *Pid*:

Anonymity: Given a pseudonym *Pid* and two identity instances:

$(Id = x_1)$
$(Id = x_2)$

where one and only one instance is correct, then it should be computationally hard to decide which is correct.

Verifiability: For each *Id* there exists a witness number *w* such that it is easy to compute *Id* and the corresponding *Pid* given *w*.

Independence: If one pair of (*Id*,*w*) is known, then the anonymity property must be true for other values of *Id*.

Note that if one knows a single witness number *w*, it is easy to compute a single pseudonym, but not more. The *w* is the 'witness' on the correct correspondence between *Id* and *Pid*, but *w* is not of any help in computing other pairs of *Id* and *Pid*. The witness number *w* enables the holder associated with *Id* to compute his pseudonym *Pid*, and it enables the holder to prove that the pseudonym *Pid* corresponds to his identity *Id*, but only for his own identity.

It must be hard to compute the witness number from a given identity *Id*, or else it would be easy to find the correspondence between *Id* and *Pid* given the requirement of verifiability. This will break the anonymity requirement. A parallel argument will be valid for computing a witness number starting from a given *Pid*.

Therefore it is hard to find both one's own witness number and others' witness numbers. Assume then that the correct witness number is given to the correct person and that this is kept confidential. Further assume that there exists a method of *testing* whether a person *knows* a witness number that corresponds to a given pseudonym without revealing the witness number. Assuming that a person only knows his own witness number and it is hard to compute a witness number, then it is possible for this test method to decide whether the person claims the correct pseudonym. This can be employed in an *anonymous verification* of claimed pseudonym when personal health record inspection is requested.

But someone must be able to compute the witness number from the identity information. A simple solution is to restrict information that must be input in this computation to some issuing authority. But how can a single party be enabled to compute something that no one else can? We have to introduce two cryptographic functions to answer this question.

One proposal that satisfies all three requirements above can be constructed with the help of oneway functions and trapdoor oneway functions. A *oneway function* is a mathematical function where it is easy to compute $g(x)$ given x, but given y it is computationally hard to find a value x such that $y = g(x)$.

A *trapdoor oneway function* is a oneway function where there exists some extra information (the trapdoor) that can be employed so that it is easy to compute the inverse function. The trapdoor information is hard to compute given the oneway function description.

Assume that f is a trapdoor oneway function and that g is a oneway function. A transformation T that takes *Id* as argument and computes a pseudonym is defined as:

$$T = g \circ f^{-1}$$

such that

$$w = f^{-1}(Id)$$

and

$$Pid = g(w).$$

The functions f and g do not need to be kept secret, but must be available both for the patient's *H* and the hospital's *TPT* to test that the witness number corresponds computationally with both *Id* and *Pid*. The function f^{-1} is restricted to the issuing authority though.

5 Registration and Verification Control

5.1 Trusted Issuer

A trusted authority *TC* will issue the witness number *w* based on *Id*. Only *TC* can do this because only *TC* knows the trapdoor function that maps from *Id* to *w*. More precisely, the computation of this function is embedded in tamper-resistant electronics so that the secret trapdoor information will not be revealed to anybody.

The *TC* must not be operated by non-authorized personnel, and the security procedures of this witness issuing function can be strengthened by several mechanisms. For instance, several *TC* operating personnel must be present when the issuing takes place. This can be done by *secret sharing* schemes, where the trapdoor information is partioned and shared among operators. The mathematical computation will take place only when all participants are present and supply their share of a secret key.

The issuing of the witness number may take place by requiring the person to visit some office. The patient's helper device *H* will receive and store the *w* automatically. The *H* will also check the correctness of the issued witness number, as described in the previous section. Now the person is capable of proving

the association with *Id* and *Pid*, by running a cryptographic protocol on his or her *H*.

5.2 Pseudonymization Facilitated by Patient

A registration record from *JR* to *LR* or *SR* is sent via the tamper-resistant pseudonym tester *TPT* for pseudonymization. The *TPT* receives the pseudonym from the patient's *H*.

One or probably more *TPT*s are installed in a hospital. The program and electronic functionality of *TPT*, including the functions *f* and *g*, are embedded in tamper-resistant electronics so that correct and complete operations can be trusted. Observe that *TPT does not* contain any secret information before and after use, but only implementation of algorithms that are assumed to be publicly available anyway. It is foremost the integrity of *TPT* that must be trusted. A *TPT* will (somewhat similar to the function of a notarius publicus) confirm that the pseudonym submitted by the patient to the *TPT* is correct with respect to the identity claimed for the actual registration record.

The patient lets *H* send *Id* and *w* to *TPT*. The *TPT* computes the identity *Id* from *w*, verifies that $Id \stackrel{?}{=} f(w)$, and checks that *Id* matches with the identity indicated by *JR*. If all this checks, then the pseudonym is computed $Pid = g(w)$, and the identity code is swapped with the pseudonym code in the registration record. The message is now ready for submission to *LR*, and the *TPT* will *delete all* computation data from its computer memory.

Note that *TPT* cannot compute the pseudonym based on *Id* only. The corresponding *w* must be supplied. Nor can the *TPT* be used to compute identities from pseudonyms. This means that the *TPT* unit can only be used to *verify* cryptographic computations already carried out. The *TPT* can be utilized only when the computed answer *w* is known in advance.

In spite of these security properties, the system managers of the hospital computer system could easily make a little program that accumulated a table containing records of (*Id*, *Pid*) *as the pseudonyms are computed*. If the security assessment of the total system organization dictates that this will pose a risk, this problem can be solved by *TPT* computing a *hospital pseudonym*, where this hospital pseudonym can only be transformed to a register pseudonym by *LR*.

Evidently, not all patients will have the possibility, or even the motivation, to own and handle an *H* that will control the registrations. It is therefore necessary that *JR* can choose to send the registration records of this category of patients to *TC*. The prerequisite is that *TC* be authorized, directly or indirectly, by the patient to carry out this transformation. A balanced security can be adopted if the *TC* can provide online simulation of the *H*, so that the protocol of pseudonymization is now happening between *TC* and *TPT* (the details of this protocol are left out here).

5.3 Anonymous Register Inspection

The patient should have the opportunity, at any time, to request and validate the registrations particular to him or her. This is facilitated by some terminal connected to the inspection service. This could be part of the functionality of a public kiosk terminal, a desktop or mobile network service. Let this terminal be denoted *VT*.

The inspection terminal *VT* will take commands from *H*. The *H* device will send a name to *VT* asking to connect to the selected registry. After connection is established, the inspection server and the *H* can engage in the inspection protocol directly. The device *H* will send *Pid*, and the server will request a proof that *H* knows the witness number. This can be carried out with a *zero knowledge protocol*, where the inspection server is convinced of this fact without *H* never having to reveal any bits of information about its *w*.

Subsequently, the inspection service will request the records indexed on the actual pseudonym from the registry, and forward the recorded data to the *H*. The person may choose between reading the information on a *VT* screen, or store the information and scrutinize the data at a later point in time at some private location. The communication channel should have end-to-end confidentiality property.

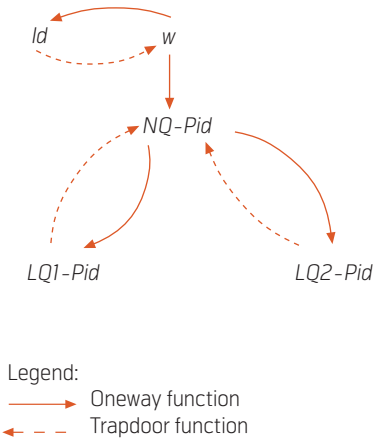


Figure 3 Registry specific pseudonyms, where functional computational dependencies are introduced between Id, w, and pseudonyms for NQ and LQ

ISSN 0085-7130

If correction is needed, the person will have to correspond with *JR*. This procedure will ensure that both *JR* and *LR* are updated and consistent.

5.4 Re-identification by Notification

Two parties in the system are given the ability to compute the identity from a given pseudonym. The patient knows her own pseudonym, through her *H*. The issuing authority *TC* can compute all pseudonyms from given identities, i.e. in principle the reverse link from a given pseudonym to the identity.

In this emerging age of ubiquitous data communications, it will soon be practical and efficient to notify the patient by broadcast email, even for the population of a country[3)]. This email will request the patient to acknowledge the pseudonym and provide an informed consent of re-identification. The procedure includes a request asking the person addressed by pseudonym to initiate further contact with an authority within some time period. If the time expires without response, the privacy policy could open for re-identification by the authority *TC*, at least in serious cases of epidemiological emergency.

5.5 Local Pseudonyms

The system can be augmented to make the identity-to-pseudonym mapping dependent on each local *LR*. A *TPG* unit must be able to transform from identity to a pseudonym for a selected *LR* or *SR*. By this we achieve restriction mechanisms against linking data stored in distinct registries. The linking ability can be assigned to one or more of the *NQ*. Then only the query processors will be able to join data from different *LR* and *SR* on pseudonym. See Figure 3.

Perhaps it is important here to emphasize that it will be a simple operation for an *NQ* unit to map from one local pseudonym to another. A practical solution will be to connect transformation to an *LR* address. Only one pseudonym type *NQ-Pid* will be internally in *NQ*.

6 Transition to Patient Controlled Registration

6.1 Registration Controlled by Trusted Third Party

A standard and general personal computing device is *widely* worn today: the wireless mobile ('PDA-phone') is rapidly becoming a serious contender for such a device. The patient will not be able to bring the cryptographic parameters and communicate with *TPT* without an electronic device *H*. This fact means that a third party must perform the pseudonymization on behalf of the patient, because we have already assumed that neither *LR* nor *JR* should be able to do this computation.

In the proposed solution with patient controlled registration, the *TC* may compute on behalf of the patients who either do not take an interest in privacy or have acute medical disabilities and have not granted others the authorization to act on their behalf. Nevertheless, by the current outlook, it is very reasonable to expect that the availability of mobile implements that can take on the functionality of *H* will increase significantly, so that the major part of the population, say more than 95 percent[4)] of the population will be able to participate in the system. This suggests a transitional phase where *all* pseudonymizations will be performed by one or more *TC* in the beginning. Then as the personal wearable computing devices become mainstream, an increasing number of patients will possess the functionality that is required from *H* in this system. An increasing number of people will be able to control the pseudonym themselves by directly communicating with the hospital's *TPT*. Observe that this implies that the main structure of the information system can stay the same from introduction to fully productional system.

If so required, the *TC* can be prevented from having read access to the health records by sending encrypted data from *JR* to *LR* or *SR*. Most likely, this will have to be performed anyway if the communication between *JR* and *TC*, and between *TC* and *LR* or *SR* will be on open networks. We require that the transport communication protocol provide authenticity and confidentiality services.

One advantage of letting *TC* carry out the pseudonymization computation is that the secret function f^{-1} *will not* be distributed and physically secured in a multitude of devices. One possible disadvantage will be that the system is now operationally dependent on online access to *TC*. In other words, if *TC* fails to be operational then no registration can be executed.

The hospital's *JR* and the local registry *LR* could be located in the geographical vicinity, and probably both units will be connected to a local or metropolitan

---

3) *This was originally proposed in 1991, the breakthrough of ubiquitous Internet and mobile communications technology has strengthened this assertion considerably.*

4) *The number of GSM subscriptions in the Norwegian population reached 4,716,090 at the end of 2004, whereas the Norwegian population was 4,640,219 at the beginning of 2006.*

area network. It follows that the communication costs will be negligible. Some costs must be attached to the registration communication with *TC*, but that will decrease with the availability of patients' *H*.

One way of avoiding the disadvantages mentioned above is to *distribute* the pseudonymization processing of *TC* to local tamper-resistant pseudonym generators *TPG*. A *TPG* unit will be able to generate and output the pseudonym given an identity as input, whereas the pseudonym tester *TPT* previously employed can not. This means that *TPG* must contain the secret part of the trapdoor function f^{-1} in a tamper-resistant package, a function reserved for the trusted third party *TC* in the previous discussion.

A general problem with utilizing a *TPG* is that the device can be exploited as a *pseudonym oracle*. One approach for managing this threat is that *TPG* will not output the pseudonym directly, but outputs a hospital pseudonym that *LR* is able to transform to a register pseudonym.

In practice, the hospital's computer system administrator could restrict the access to *TPG* service. However, the result will be a weaker balanced security than in a patient controlled pseudonymization process where it is *impossible* to employ the *TPT* as a pseudonym oracle.

It is reasonable to predict that increasingly the patients will own or be given easy access to a personal computing device *H*. Hence the person, using the *H*, will be able to check what is stored in the registry about him or her immediately, and without prior arrangement with the hospital. Inherently, this will be an important incitement to purchase the *H* and receive the witness number *w*. As the majority of users hold a *H* containing the *w*, the declining demand of the service provided by the local *pseudonym generators TPG* can be centralized to a *TC* as the *pseudonym testers TPT* are replacing the *pseudonym generators*. This will shift the work load from the realtime online service of the *TC* server, and localize the data communication, hence reducing cost and improving reliability and security.

6.2 Inspection Arrangement

Envision the situation where the hospitals are using the *TPG* issued by the trusted third party *TC*. The patient does not possess a *H*. How can the right to inspect be secured in this situation?

The hospital's *TPG* can issue a 'ticket' that stores the patient's witness number. For instance, the ticket could be realized by very cheap technology of a magnet stripe card. The perceived threat that unauthorized individuals might access the ticket can be prevented by data encryption derived from some password. The password can be freely chosen by the patient, or better, the *TPG* chooses a random password and writes this in a sealed envelope, similar to current issuing technology of credit card companies of PIN. Subsequently, the ticket and the password can be used in a verification terminal *VT*.

The ticket is brought to *VT*. The terminal *VT* must be assumed to operate on behalf of the person and be trusted not communicate or store identities and health records. The *VT* will read the ticket, ask for the password, decrypt the witness number, and compute the pseudonym $Pid = g(w)$. Subsequently, the *VT* connects to the registry, and requests the information stored under *Pid*. The registry's inspection program will ask for authentication, quite similar to the patient controlled registration. If the authentication is verified the health data will be returned to *VT*, which forwards the information to the patient by screen or printout. Finally, the user will log out and the terminal will delete all data from the transaction.

Note that this inspection scheme assumes that the patient will trust that *TPG* and *VT* are correct. This and other trust assumptions are not required if the patient holds the trusted *H*. As remarked before: A system where the organization equip the client with *its own* instruments necessarily becomes securitywise unbalanced and requires strong trust demands of the client.

6.3 Re-identification

Assume that a user of *LR* is authorized to receive *Pid* from *LR* via *LQ*. For example, the user wants to target information about a new medical treatment to a special group of patients, or make some follow-up investigations of such a group. The right and procedure for re-identification within the scenario of tickets must be carried out by some third party authority. A natural choice would be the *TC*, which also should take on the task of security policy and management.

6.4 Extant Registries

Special medical databases for epidemiological purposes must be assumed to be operational already[5)]. This is accommodated for in the proposed architecture by the *SR* registries. The health records in these *SR* registries are indexed by unique personal identity numbers. This indexing must be changed to pseudonyms if these health records should be joined with individual records in a registry (*NQ* unit). This translation process can be done by the *TC* as a one-time

5) *Four specialized national health registries operated in Norway 1992.*

task. At the same time all *JR* have to start sending pseudonym records to *SR* as described.

A practical transitional procedure to a new structure would be to maintain the existing practice while starting the new pseudonym registry. The ongoing registrations will go to the established registries, as before. In addition, both old and new records are sent to the new pseudonym indexed registry by using a special *TPG* unit issued by *TC*. The pseudonym type of this *TPG* will be special to this registry. The result is that even if the *SR* organization knows the relation between identity and this special pseudonym, this knowledge will not be applicable to other registries because they employ other mappings. A central *NQ* unit will, as described before, be enabled to link the different pseudonym types.

The organizational separation between the existing identity-based registries, and the new pseudonym-based registry, similar to the distinction between *JR* and *SR*, might secure that linking cannot take place. If this can be introduced then a standard pseudonym type can be employed.

While the hospitals will carry out the transition to a pseudonym-based registration and send the registration directly to *SR*, the number of registration messages submitted to the old identity-based registry will steadily decrease until all hospitals have converted to the new system. All health records can be joined continuously from a user perspective, even in this transition phase, because all records will be available at the *SR* by pseudonyms.

7 Interactive Statistics

7.1 Client-Server

Traditionally, there have been two main methods to issue statistical data to users: macro and micro statistics [5].

Macro statistics
This is a collection of related accumulations, normally presented by two dimensional tables. The disadvantage with this is that the revealed data can only present a very limited basis for statistical analyses.

Micro statistics
One or several files containing individual database records are stored on some storage medium and transferred to the user. The user will utilize his own computer program to perform statistical analyses. The database issuer cannot restrict the functionality of these programs in any way (they can be constructed by the user). Therefore the protection against dissemination of sensitive information must be done in advance. For instance, this could be removal of name fields, records with extreme values, extract a randomized subset, or add 'noise' to resulting values.

None of the methods described above exploit the advantages of a state-of-the-art distributed interactive database system, where the users are online to the database employing the tools of an application-oriented query language. The system architecture proposed here builds upon the *client-server model*. A user's workstation will act as a client. A coordinating query processor will be a server to the workstation, but will take the role of a client with respect to one or more database machines. A database machine will normally take a server role. A client machine will be able to submit queries over the data communications network to many database server machines, then collect and process the response on the local machine.

The researcher is equipped with a workstation that runs program tools for statistical analyses and as front end clients to database systems, and network communication program services that connect to the statistical services. Hence the user's geographical location is independent of the registry's location.

Every local statistical registry provides a networked service. Access control is performed by authentication of the query messages, for example determined by client identity and query type. When access is permitted, the query is forwarded to a server application (*LQ*) which performs the detailed database transactions.

The response from the database system will be sent to the server application as a database table. The server application will carry out the requested computations (statistical functions) on the table. When this computation is finished, the result will be conditionally returned from *LQ* over the communication network to the work-station that generated the query. The result is allowed to be sent if it checks against the inference security policy of *NQ*. For instance, the statistics should not release unnecessary pieces of detailed information of the individual patient records to the researcher or administrator.

7.2 Statistical Functions

Let us review the basic statistical functions that should be available in a statistical database server, under the assumption that the work-stations used by the researcher can carry out extensive statistical processing of estimators and testing of hypotheses.

Let us start out with a small example. Assume a simple database table where each record (row in the table) contains values for the following five attributes (synonymous with fields or columns in the table):

Identity number Date of birth Gender Date Diagnosis

The code in the field *Identity number* will be transformed to another code denoted a pseudonym when communicated from *JR* to *LR*. It follows that the corresponding table in *LR* will contain records of the following structure:

Pseudonym number Date of birth Gender Date Diagnosis

Assume further that this table in *LR* contains *N* records. A subset of the records can be characterized by a predicate *C*. As an example:

C: (*Gender* = `male`) AND
(*Diagnosis* = `arthritis` OR `hypertension`).

The table will be partitioned in two, the part that satisfies *C* and the part that does not satisfy *C*. The size of the part that satisfies *C* will be a variable dependent on the instance of predicate expression C and the actual records stored in the *LR* table.

Assume a general registry table, where the fields are numbered sequentially from left to right 1, 2, 3, ..., *j*, …, and where the j^{th} field represents that attribute A_j. The statistical counting is computed over the partition of the table characterized by the expression *C*.

The simplest statistical functions are *count*, *frequency* and *sum* (for attributes where addition is defined). The functions *Relative frequency* and *Average* are computed:

$$\mathbf{rfreq}(C) = \frac{\mathbf{count}(C)}{N}$$

The function **count** can be employed in the example given above by simply counting the number of records that satisfies the values assigned to the attributes *Gender* and *Diagnosis*.

One example of the application of the functions **sum** and **count** is: We want to find the average of *systolic pressure* (assuming this is an attribute in the table). First all the blood pressure values must be summed up over all the records that satisfy *C*, then the total number of such record instances is counted:

$$\mathbf{avg}(C, A_j) = \frac{\mathbf{sum}(C, A_j)}{\mathbf{count}(C)}$$

More general statistical computations are *variance*, *covariance* and *correlation-coefficient*. This and higher order moments and central-moments can be computed from a general query in accordance with the form

$$\mathbf{q}(C; e_1, e_2, ...) = \sum_{i \in C} x_{i_1}^{e_1} x_{i_2}^{e_2} ...,$$

where the exponents e_1, e_2, … are non-negative integers, and x_{i_1} is the value of A_1 in row *i* in the table.

Other types of statistics that could be relevant are *median*, *percentile*, and extremal points *max*, *min*.

8 Users

8.1 Security Policy

The preceding discussion has introduced the notions of *inference policy and control*. This can be made more precise now, by stating that the inference policy is a set of rules that can decide which types of statistics a user is authorized to have access to, and which pieces or information are *sensitive*, meaning that the computed response to the query will release personal information that breaks the policy rules.

The security measures to be installed will depend on the detailed security policy determined. The proposed architecture has great flexibility here. It is possible to differentiate between local and special registry, between users, between access rights of the various *NQ* to data, and so on. In most cases it will be possible to keep the statistical functions completely local so that *NQ* does not have to manage raw data. This means that users who do not have access rights to raw data still are able to request useful *NQ* functionality.

Recall that the structural starting point of the architecture was the introduction of a strong separation of data source and data sink, as shown in Figure 1. The input of personal health data takes place on one side, while the statistical output is carried out on the other side. If a person is authorized to carry out both registration of health data *and* statistical queries using *NQ*, this person will in principle be able to link identity and pseudonym easily. This is so because the person is able to modify the database in a controlled way, for instance data perturbations that may easily be detected and identified. This and similar problems belong to the problems of inference control of data base systems, which is now developed further.

8.2 Modes of Operation

Fundamentally, there appears to be three different work mode possibilities for a research scientist in this respect.

1 The health management researcher Alice will connect her work-station to one or several of the registries *LR* and *SR*, and make statistical queries of the type described in the previous section. (Of course, other types of queries can be done too, but are not analyzed here.) The researcher receives numbers concerning counts, summations and correlation coefficients of attribute variables conditioned on a given characterization. Further statistical analysis and testing will take place locally on the work-station, using her favourite software client tools for this purpose.

2 The clinical researcher Bob has acquired his own database that can be processed locally on his work-station with accompanying tools. Access control and information flow are at the researcher's discretion.

3 The epidemiological researcher Cecilie acquires her own statistical database and *links* this to extant institutional medical registries. Basically, two distinct motivations might be found for this linking requirement:

   a Either the need to link emerges *after* the data is collected,

   b Or a selected subpopulation of the registry forms the baseline for collecting more medical data about these persons.

The implications of modality 3 are taken further in the next Section.

8.3 Linking

8.3.1 Join

The operation **join** is well defined in a relational database. Linking two or more tables in this context is limited to **equijoin** over the identity attribute (the pseudonym).

Let *the researcher's table* comprise an index and two attributes,

$$T_F = \{(\#, Id, A_F)\}.$$

Moreover, let *the registry table* be

$$T_R = \{(\#, Pid, A_R)\}.$$

The object of employing the function **equijoin** is to form a table of tuples (Pid, A_F, A_R) where the value A_F belongs to the correctly derived *Id* from *Pid*. The next section will discuss how this can be done.

8.3.2 Linking Research and Institutional Data

In this case the researcher has already collected the health data indexed on identity, and wishes ad hoc to align these data with some records of the registry. In this section, *Q* will denote either *LQ* or *NQ*, and *R* will denote one or more *LR*. Two possible alternatives emerge here.

1 The hospital performs the registration of data T_F and sends this further to *LR*, either via the hospital's *TPT* if the patient is there, or else by the *TC*. The researcher is authorized to acquire statistics via *Q*. This solution becomes very similar to what is sketched already for information sent from *JR* to *LR*.

2 Let T_F be pseudonymized, either by an assigned research *TPG* (for instance in the form of a smart card) or by submitting data to *TC*, which will transform from $Id \rightarrow F\text{-}Pid$ (researcher pseudonym). This transformation will be according to the same principle as described for the local registry pseudonym. Similarly, *Q* must be enabled to transform from *F-Pid* to *Q-Pid* (registry pseudonym) when the pseudonymized table is sent to *R*. This solution is independent of the hospital's computers. The method *does not* release the pseudonyms of the registry, which will remain an internal index. Instead the researcher is given a pseudonym that can be transformed internally by *Q* into a registry pseudonym. Note the analogy between *LQ* with respect to a researcher pseudonym, and *NQ* with respect to a local registry pseudonym (Section 5.5).

Linking: First, let us give the details of alternative 2 described above. The researcher will receive a table transformed from (Id, A_F) to $(F\text{-}Pid, A_F)$, either with the help of a *TPG* issued by *TC*, or the *TC* performing this directly.

A straightforward procedure is that the researcher sends the pseudonym-indexed table to *Q*, which computes *Q-Pid* from *F-Pid*. Now *Q* is able to compute **equijoin** which results in the table

$$T_{FR} = \{(Q\text{-}Pid, A_F, A_R)\}.$$

Thereafter, the researcher can perform the statistical investigations aided by *Q* on T_{FR}.

A more complicated, but storage-efficient way would be that *Q* generates a *temporary* T_{FR} dependent on the researcher's queries. Five different cases emerge, depending on where the characteristic attributes are located and which attributes the query *q* contains. In general, the researcher will submit a subset of all researcher pseudonyms to *Q*, possibly together with

the attributes necessary. The actual linking will be performed by *Q*, returning the statistical results.

8.4 From Pseudonym to Identity

Consider the case of transforming from pseudonym to identity. The baseline identified for a closer monitoring will be a subpopulation of *R*. This requires *re-identification*, something we have made great strides to make impossible! In a controlled manner, this can be done by *TC* if *Q* on request supplies the pseudonyms of the subpopulation and the researcher's identity to *TC*. The *TC* makes a decision to issue a list of identities to the researcher.

Note that the researcher *does not* receive the pseudonyms. The researcher collects new medical information indexed on *Id*. If the new data should be linked to *R*, we end up in a situation as described in Section 8.3.2.

Authorization by the Patient. If we may assume that patients possess *H* and so are capable of computing the pseudonym, new possibilities open up with a view to continuing health investigations of patients. A patient can compute a research pseudonym *Fid*, and by this computation directly acknowledge that collected information can be linked to his or her medical data in *R*. Hence the linking can be directly authorized by the patient, so person identification becomes irrelevant with respect to the health data acquisition itself.

8.5 The National Level

The starting point of the preceding discussion was the application of a single or several non-connected statistical registries. A cohesive national level health registry can be achieved by joining all local statistical registries into a distributed system that provides the statistical researcher with a coordinating query-processor *NQ* service. The *NQ* entity will provide national level query services by coordinating communication with all the local registries.

The *NQ* query processor can be implemented by a server computer that authorized users are able to access over the computer network, independent of user-location. Alternatively, the *NQ* functionality can be realized by a software process on the user's workstation. Accesses restrictions to the local registries are controlled by the *LQ* entities.

Under the assumption that all *LR* can provide the same datastructure or schema, and no duplication exists, then the statistical functions of a query can be computed at each *LQ* and be combined in *NQ*. Only aggregated information will be available in *NQ*.

This implies that it may be possible for the client (the user's workstation) to perform the *NQ* query processing. The client sends out queries to all available *LQ* directly, each *LQ* returns the response, and the client collects and combines the responses into a national level query response.

However, legacy systems may create problems with respect to interoperability. In general, we must be prepared to allow for the local datastructure (schema) to be distinct for two or more local registries, for instance, each special registry *SR* will have a distinct datastructure. This makes it more complicated, but the analysis will be similar to the integration of a researcher's special registry and a common statistical registry, as discussed in Section 8.3.2.

The easy solution in linking a special registry and a common health registry is simply to send the special registry to the *LQ* or *NQ*. But this goes against the requirement of not transmitting 'raw data' from one *LQ* to another *LQ*.

If a covariance computation between attributes located on *different* local registries is *not* required, then a query can be based on local computations by each *LQ* acquired by the *NQ*. Only aggregated data need to be sent to *NQ*. Hence, this type of query can be performed from a researcher's work-station.

If a covariance or similar computation on attributes located on *different* registries is required, then both the pseudonym and the relevant attribute values must be sent from *LR* to *NQ* for temporary use. The data can be deleted when *NQ* has performed the statistical computations. Monitoring that *NQ* does not build its own database of received records must be carried out by administrative procedures.

The characterizing attributes are the fields in the database that is used to describe the subpopulation that is under investigation. The count attributes are the fields in the database which the statistics are based on. Often, but not necessarily, the characterizing attributes are identical to the count attributes.

If the characterizing attributes and the count attributes in a query are located on distinct registries, a list of pseudonyms can be sent from one *LR* to the other, coordinated via the *NQ*. If different registry pseudonyms are used in these registries, the coordinating machine must do the translation. The recipient *LQ* will use this pseudonym-list as part of the query characterization, and subsequently delete this when the result of the computation is obtained.

The local registries *LR* will not be able to link records if the pseudonym type is dependent on the *LR* instance. A coordinating machine *NQ* must be admitted to have access to the various registry pseudonyms and their translation, and even the attributes under the special requirements discussed above. Normally, the users will receive only the evaluation of statistical functions requested. Still, it is possible to admit access to all 'raw data' in *LR*. For instance, this could be useful in connection with quality assurance of data registered.

9 Feasibility

The proposed system can be based on readily available technology. The following is an assessment in this respect.

The *cryptographic functions* applied can be constructed from public-key cryptography. Several candidate functions exist, and a more detailed evaluation must be carried out taking into account implementation constraints.

Tamper-resistant hardware technology is available in the form factor of smart cards, PCI-bus boards, and stand-alone computing machinery. The exact physical protection required must be based on a risk assessment of the operations.

Data communication requirements are trivially satisfied by Internet communication technology.

Distributed middleware: If existing *SR* shall be included, the system will represent a heterogeneous multi-database system; that is, the local database systems are not uniformly constructed with respect to query language, data structure and coding. Of course, it will be a much more difficult challenge to include legacy systems than to build a homogenous distributed database system top-down. The contemporary practice of three-tiered systems and mobile code are promising components toward a general solution to this challenge. Moreover, the proposed distributed database system is simpler than a general online interactive system, because the distributed application is limited to 'reading' from the database, whereas the 'writing' to an *LR* will be carried out by a single source: *JR*.

10 Conclusions

This paper has introduced and described a multiparty security architecture protecting the privacy of medical records in statistical databases. The system design alleviates the ambivalence between data keyed to person identity and the need of epidemiologic research. Each individual controls correct and complete registration and is enabled to inspect his or her records anonymously. Public health administrations can collect records of patients systematically, providing controlled online server access to a variety of health statistics without releasing individual attributes. Mechanisms and services for follow-up studies and patient-controlled re-identification are supported. The solution scales well because of its distributed and localized approach using multiple autonomous database management systems. Furthermore, legacy systems are included as part of the system introduction.

Additional work is needed to gain experimental validation and insights into the proposed design. One interesting direction would be to attempt implementation based on a three-tier Web architecture. Also, further research into more general technical security challenges raised by the system design is necessary, including constructing access control policies, analyzing statistical query processing and optimization related to inference control, and constructing secure multiparty protocols that can replace functionality now put on trusted third parties.

Acknowledgments

Once upon a time, the committee's secretary Cand.Jure. Ingvild Mestad offered me the challenge of supplying a technical contribution to the discussions of the committee and meet. I wish to thank the enthusiastic committee leader professor Dr.Jure. Erik Boe. Thanks for mutual inspiration and cooperation across our scholarly borders. Thanks also to my friend Dr.Ing. Kenneth Iversen for sharing his insight into health informatics.

References

1 Boe, E. Dept. of public and international law, Univ. of Oslo. *Personal communications*, 1991-1992.

2 NOU 1993:22, vedlegg 1. *Pseudonyme helseregistre*. Oslo, Sosialdepartementet, 1993, 281-305. (ISBN 82-583-0360-0)

3 *Lov om helseregistre og elektronisk behandlling av helseopplysninger*. Oslo, Helse- og Omsorgsdepartementet, Dec 1998. Available from: http://www.odin.dep.no/shd/norsk/publ/hoeringsnotater /030005-990298/index-dok000-b-n-a.html. (January 2001)

4 Garfinkel, S. *Database Nation. The Death of Privacy in the 21st Century*. O'Reilly, 2000.

5 Denning, D. *Cryptography and Data Security*. Addison-Wesley,1982.

6 Lunt, T F (ed). *Research Directions in Database Security*. Springer, 1992.

7 Tamer Özsu, M, Valduriez, P. *Principles of Distributed Database Systems*. Prentice Hall, 1999.

8 Brandt, J, Damgård, I, Landrock, P. Anonymous and verifiable registration in databases. *Proc. of Eurocrypt'88*, 166–176.

9 Mjølsnes, S F. Privacy, Cryptographic Pseudonyms and The State of Health. *Advances in Cryptology – ASIACRYPT'91*. (Lecture Notes in Computer Science, 739.) Springer, 1993.

10 Chaum, D. Group Signatures. *Advances in Cryptology – Eurocrypt' 91 Proceedings*. Springer, 1991, 257-265.

11 Mjølsnes, S F. *Health Records by Cryptographic Pseudonyms*. Invited talk at VLDB Workshop on Secure Data Management, SDM'05, Trondheim, 2 September 2005.

Abbreviations

TPG	Pseudonym generator based on tamper-resistant microcontroller
TPT	Pseudonym tester based on tamper-resistant microcontroller
H	Patient's digital helper; a pocket computer such as a smart card
JR	Hospital patient journal roster
LQ	Query processor associated with *LR*
LR	Local research and administrative medical registry
MIX	A network node that hides the original sender address
NQ	A central query processor
SR	Special constructed health registry
TC	A reliable, trusted central computer/authority
VT	A publicly accessible terminal for inspection of personal data
A	Health attributes
Id	Identity string, for instance social security number
Pid	Pseudonym number
w	Witness number for identity and pseudonym

Stig Frode Mjølsnes received the Siv.Ing. degree in Physical Electronics in 1980 and the Dr.Ing. degree in Telecommunications in 1990, both at the Norwegian Institute of Technology, Trondheim. The doctoral thesis was in the field of cryptographic protocols. Main affiliation from 1983 through 1999 was the contract research organization SINTEF as a research scientist, working primarily on communications security and applied cryptography projects. He lectured as adjunct Associate Professor at the Department of Computer Engineering, NTNU through 1999–2001. During the period 2000–2002 he held the position of Associate Professor at the University of Stavanger. From 2003 he holds a full professorship in information security at Department of Telematics, NTNU. He is committee executive manager of NTNU Research Programme for Information Security under the strategic focus area of ICT at NTNU, and leads the information research group at the department.

email: sfm@item.ntnu.no

Resources

VLADIMIR A. OLESHCHUK, GEIR M. KØIEN

Web Pages to Visit

The following is a random list of web based resources on personal privacy. It ranges from the technically oriented to sites concerned with democracy and legal rights. The sources cover everything from easy-to-read light entertainment accounts of user privacy issues to more profound academic papers. The homepage sources should also provide you with a good starting point for your own explorations in the field, be it for 'information' or for more serious study.

Note that the inclusion of any of these references does not necessarily mean that we endorse the contents on the homepages or that we vouch for the correctness or quality of the material contained on the homepage.

Center for Democracy and Technology Web reference: http://www.cdt.org/	"The Center for Democracy and Technology works to promote democratic values and constitutional liberties in the digital age. With expertise in law, technology, and policy, CDT seeks practical solutions to enhance free expression and privacy in global communications technologies. CDT is dedicated to building consensus among all parties interested in the future of the Internet and other new communications media."
The Free Haven Project Web reference: http://freehaven.net/	The Free Haven project began in December 1999 as a research project initially comprised of several MIT students to design, implement, and deploy a functional data haven. The project is privacy research oriented, but also aims at producing practical solutions.
The Onion Router (TOR) Web reference: http://tor.freehaven.net/ or http://tor.eff.org/	"Tor is a network of virtual tunnels that allows people and groups to improve their privacy and security on the Internet. It also enables software developers to create new communication tools with built-in privacy features. Tor provides the foundation for a range of applications that allow organizations and individuals to share information over public networks without compromising their privacy."
U.S Navy Onion Routing web page Web reference: http://www.onion-router.net/	This homepage provides an overview of the Onion Routing research program. This program is made up of projects on researching, designing, building, and analyzing anonymous communications systems. The homepage is an official US navy web site operated by US Naval Research Laboratory (http://www.nrl.navy.mil/).
Electronic Frontier Foundation (EFF) Web reference: http://www.eff.org/Privacy/	"EFF is a nonprofit group of passionate people – lawyers, technologists, volunteers, and visionaries – working to protect your digital rights." On Privacy: "New technologies are radically advancing our freedoms, but they are also enabling unparalleled invasions of privacy."
Microsoft – Privacy Guidelines for Developing Software Products and Services Web reference: http://www.microsoft.com/downloads/details.aspx?FamilyID=c48cf80f-6e87-48f5-83ec-a18d1ad2fc1f&displaylang=en	This is a Microsoft guideline document for developers. "Brief Description: This document is a set of privacy guidelines for developing software products and services that are based on our internal guidelines and our experience incorporating privacy into the development process."
Microsoft – I Know What You Did Last Logon – Monitoring Software, Spyware, and Privacy Web reference: http://www.microsoft.com/downloads/details.aspx?FamilyID=c48cf80f-6e87-48f5-83ec-a18d1ad2fc1f&displaylang=en	This is a Microsoft whitepaper. It explores ... "the technical methods employed by both hardware and software-based key loggers, how keystroke loggers are integrated with specific malware threats, the user experience associated with various key loggers installed, and examine the social and legal appropriateness of various use scenarios ..."
Wikipedia – Spyware Web reference: http://en.wikipedia.org/wiki/Spyware	This is the Wikipedia homepage on Spyware. *It is noted that Wikipedia is a source of information that must be used with caution as the editorial policies of Wikipedia do not meet the standards of normal editorial policies.*

Wikipedia – Privacy Web reference: http://en.wikipedia.org/wiki/Privacy	This is the Wikipedia homepage on Spyware. *It is noted that Wikipedia is a source of information that must be used with caution as the editorial policies of Wikipedia do not meet the standards of normal editorial policies.*
Privacy.org Web reference: http://www.privacy.org/	This is an advocacy webpage. The slogan is "Privacy is a Right, not a preference". Contains many articles on privacy and useful links.
IBM Privacy Research Institute Web reference: http://www.research.ibm.com/privacy/	Mission Statement: "The IBM Privacy Research Institute is an organization within IBM Research to promote and advance research in privacy and data protection technology. Our goal is to develop technologies for enterprises to conduct e-business in privacy-enabling ways. The institute's research focuses on technologies for commercial applications, particularly for e-business."
W3C – Platform for Privacy Preferences (P3P) Project Web reference: http://www.w3.org/P3P/	This World Wide Web Consortium (W3C) project has now been suspended, but the homepage still contains useful information.
W3C Workshop on Languages for Privacy Policy Negotiation and Semantics-Driven Enforcement Web reference: http://www.w3.org/2006/07/privacy-ws/report	This is the workshop report homepage for the W3C Workshop on Languages for Privacy Policy Negotiation and Semantics-Driven Enforcement, Ispra, Italy, 17-18 October 2006. The page contains a lot of useful material.
Privacy Enhancing Technologies Web reference: http://petworkshop.org/	This is the homepage for the annual Privacy Enhancing Technologies Workshop. In addition to information about the current workshop the site also contains information on the previous workshops and an anonymity paper reference library (at http://petworkshop.org/2007/links.php).
Office of the Privacy Commissioner of Canada (OPC) Web reference: http://www.privcom.gc.ca/index_e.asp	Canada seems to be paying a lot more attention to privacy than most nations. The resource center (http://www.privcom.gc.ca/information/index_e.asp) contains a lot of useful information, publications and references. Recommended!
Office of the Information and Privacy Commissioner of Ontario (IPC) Web reference: http://www.ipc.on.ca/	They do insist on privacy in Canada. Again there is a lot of information available. The site covers both research topics and provides a lot of useful links. Recommended!
RSA Laboratories Web reference: http://www.rsa.com/rsalabs/	RSA Laboratories have a homepage dedicated to RFID Privacy and Security at http://www.rsa.com/rsalabs/node.asp?id=2115 . There are many RSA Labs papers available at this homepage.
RFID Consortium for Security and Privacy Web reference: http://www.rfid-cusp.org/	"RFID CUSP is a partnership between academic and industrial scientists specializing in RFID security and privacy. Our mission is to make RFID safe for consumers by conducting open research and educating the next generation work force that will develop, deploy and maintain secure RFID infrastructures." The homepage contains a fair amount of useful information. A number of interesting papers are available/referred to at http://www.rfid-cusp.org /publication.html
Electronic Privacy Information Center (EPIC) Web reference: http://www.epic.org/	"EPIC is a public interest research center in Washington, DC. It was established in 1994 to focus public attention on emerging civil liberties issues and to protect privacy, the First Amendment, and constitutional values." The homepage contains a fair amount of useful information, including references to privacy tools (http://www.epic.org/privacy/tools.html) and other resources (http://www.epic.org/privacy/privacy_resources_faq.html).
Crypto-Gram Newsletter (Bruce Schneier) Web reference: http://www.schneier.com/index.html	Bruce Schneier is a technologist and author. In the post 9/11 era he has focused more and more on privacy in his regular Crypto-Gram Newsletter. Subscription (free) at http://www.schneier.com/crypto-gram-sub.html. If you don't feel like providing your email address then the contents is available at http://www.schneier.com/crypto-gram.html. Privacy related articles at http://www.schneier.com/cgi-bin /search/search.pl?Realm=whole+site&Terms=Privacy

	The homepage contains a fair amount of useful information, including references to privacy tools (http://www.epic.org/privacy/tools.html) and other resources (http://www.epic.org/privacy/privacy_resources_faq.html).
Anonymity terminology paper Web reference: http://dud.inf.tu-dresden.de /Anon_Terminology.shtml	This is a "consolidated terminology" paper updated by Andreas Pfitzmann and Marit Hansen. They "propose a terminology which is both expressive and precise. More particularly, we define anonymity, unlinkability, unobservability, pseudonymity (pseudonyms and digital pseudonyms, and their attributes), and identity management." The paper is available at *TU Dresden* on a homepage at the *Factulty of Computer Science*. The homepage with the anonymity terminology paper is hosted on the *Privacy and Security publications* homepage (http://www.inf.tu-dresden.de /index.php?node_id=703). This homepage contains many privacy related papers (many in German only).
The Register (Security) Web reference: http://www.theregister.co.uk/security/	The Register is an IT news site produced by Situation Publishing Ltd, London, UK. They regularly feature articles on privacy, and more often than not taking a strong stand. Amongst the articles you'll find here are the likes of *Would you trade your password for chocolate?* (the answer, apparently, is that a good many of us would; 70% was the cited number. 34% would do it without even being tempted by the chocolate); *WEP key wireless cracking made easy* (it takes less than a minute to crack WEP now); *Microsoft admits WGA update phones home* (the Windows Genuine Advantage program will phone home to Redmond even if the user clicks cancel); and *Security, privacy and DRM: My wishes for 2007* (wishful thinking).
Security and Privacy in RFID Systems Web reference: http://lasecwww.epfl.ch/~gavoine/rfid/	This homepage is dedicated to academic works on security and privacy for RFID systems. It contains a reasonably complete listing of conferences and papers in the field. Quote: "The goal of this page is to reference works related to security and privacy in RFID systems. The bibliography contains references toward refereed papers published in journals and conference proceedings, as well as technical reports and theses. It is updated on an irregular basis depending on the flow of papers published in the domain".
The Journal of Privacy Technology Web reference: http://www.jopt.org/	"The Journal of Privacy Technology is a refereed online journal published by the Institute of Software Research, a division of the School of Computer Science at Carnegie Mellon University in Pittsburgh, Pennsylvania. The Journal is a forum for publication of current research in privacy technology. It will consider any material dealing primarily with the technological aspects of privacy or with the privacy aspects of technology, which may include analysis of the interaction between policy and technology or the technological implications of legal decisions."
Data Privacy Lab Web reference: http://privacy.cs.cmu.edu/	"The Laboratory for International Data Privacy (also known as the "Data Privacy Lab") at Carnegie Mellon University is dedicated to creating technologies and related policies with provable guarantees of privacy protection while allowing society to collect and share private (or sensitive) information for many worthy purposes. We do this by partnering with institutions, agencies, and corporations facing real-world privacy concerns."
The Privacy Rights Clearinghouse Web reference: http://www.privacyrights.org/index.htm	The Privacy Rights Clearinghouse (PRC) is a nonprofit consumer organization with a two-part mission – consumer information and consumer advocacy. It was established in 1992 and is based in San Diego, California. It is primarily grant-supported and serves individuals nationwide.
HPP Web reference: http://www.healthprivacy.org/	"The Health Privacy Project (HPP) is dedicated to raising public awareness of the importance of ensuring health privacy in order to improve health care access and quality, both on an individual and a community level. The Health Privacy Project is dedicated to raising public awareness of the importance of ensuring health privacy in order to improve health care access and quality, both on an individual and a community level. The Project is a part of the Institute for Health Care Research and Policy at the Georgetown University Medical Center."
TRUSTe Web reference: http://www.truste.org/about/	TRUSTe® is an independent, nonprofit enabling trust based on privacy for personal information on the internet. It certifies and monitors web site privacy and email policies, monitor practices, and resolve thousands of consumer privacy problems every year.

Terms and Acronyms in Privacy in Telecommunications

Acronym /Term	Definition	Explanation	Web-resources
2G	Second Generation mobile technology	Refers to the family of digital cellular telephone systems standardised in the 1980s and introduced in the 1990s. They introduced digital technology and carry both voice and data conversation. CDMA, TDMA and GSM are examples of 2G mobile networks.	
3G	Third Generation mobile technology	The generic term for the next generation of wireless mobile communications networks supporting enhanced services like multimedia and video. Most commonly, 3G networks are discussed as graceful enhancements of 2G cellular standards, like e.g. GSM. The enhancements include larger bandwidth, more sophisticated compression techniques, and the inclusion of in-building systems. 3G networks will carry data at 144 kb/s, or up to 2 Mb/s from fixed locations. 3G comprises mutually incompatible standards: UMTS FDD and TDD, CDMA2000, TD-CDMA.	
3GPP	Third Generation Partnership Project	Group of the standards bodies ARIB and TTC (Japan), CCSA (People's Republic of China), ETSI (Europe), T1 (USA) and TTA (Korea). Established in 1999 with the aim to produce and maintain the specifications for a third generation mobile communications system called UMTS. Note that 3GPP is not itself a standisation organisation and that all produced standards must be ratified by a standardisation organisation. A permanent project support group called the Mobile Competence Centre (MCC) is in charge of the day-to-day running of 3GPP. The MCC is based at the ETSI headquarters in Sophia Antipolis, France.	http://www.3gpp.org
3GPP2	Third Generation Partnership Project 2	A collaborative third generation (3G) telecommunications specifications-setting project comprising North American and Asian interests developing global specifications for ANSI/TIA/EIA-41 Cellular Radiotelecommunication Intersystem Operations network evolution to 3G and global specifications for the radio transmission technologies (RTTs) supported by ANSI/TIA/EIA-41. 3GPP2 was initiated as a result of the International Telecommunication Union's (ITU) International Mobile Telecommunications IMT-2000 initiative, covering high speed, broadband, and Internet Protocol (IP)-based mobile systems featuring network-to-network interconnection, feature/service transparency, global roaming and seamless services independent of location. 3GPP2 is a collaborative effort between five officially recognised Standards Development organisations (SDO): ARIB – Association of Radio Industries and Businesses (Japan), CCSA – China Communications Standards Association (China), TIA – Telecommunications Industry Association (North America), TTA – Telecommunications Technology Association (Korea), and TTC – Telecommunications Technology Committee (Japan).	http://www.3gpp2.org
AAA	Authentication, Authorization and Accounting	Key functions to intelligently controlling access, enforcing policies, auditing usage, and providing the information necessary to do billing for services available on the Internet. The term AAA is used to denote an internet security service architetcure that provides the AAA services. The arcitecture includes AAA servers and AAA protocols. The AAA protocols include RADIUS and DIAMETER. Defined in IETF RFC 2903.	http://www.ietf.org, http://tools.ietf.org /html/rfc2903
ADSL	Asymmetric Digital Subscriber Line	A data communications technology that enables faster data transmission over copper telephone lines than a conventional modem can provide. The access utilises the 1.1 MHz band and has the possibility to offer, dependent on subscriber line length, downstream rates of up to 8 Mb/s. Upstream rates start at 64 kb/s and typically reach 256 kb/s but can go as high as 768 kb/s. Specified by ANSI T1.413 and by ITU-T recommendation G.992.1. Aversion called ADSL Lite providing up to 1.5 Mb/s downstream rates is specified as G.992.2.	http://www.itu.int
AES	Advanced Encryption Standard	Also known as Rijndael. In cryptography, it is a block cipher adopted as an encryption standard by the US government. It is expected to be used worldwide and is analysed extensively, as was the case with its predecessor, the Data Encryption Standard (DES). AES was adopted by the National Institute of Standards and Technology (NIST) in November 2001 after a 5-year standardisation process.The cipher was developed by two Belgian cryptographers, Joan Daemen and Vincent Rijmen, and submitted to the AES selection process under the name 'Rijndael', a blend comprising the names of the inventors.	http://www.nist.gov
AH	Authentication Header	AH is an IPsec protocol. This protocol is no longer needed in IPsec, but is retained for backward compatibility reasons. Defined in IETF RFC 4302.	http://www.ietf.org/, http://tools.ietf.org /html/rfc4302
AKA	Authentication and Key Agreement	A challenge-response based authentication cryptographic protocol that additionally also includes agreement on session key material. In the 3GPP sphere there exists several variants, including GSM AKA, UMTS AKA, IMS AKA etc. Note that the 3GPP2 CDMA2000 system uses an AKA protocol almost identical to the UMTS AKA protocol. Specified in 3GPP TS 33.102.	http://www.3gpp.org /ftp/Specs/html-info /33102.htm

Acronym /Term	Definition	Explanation	Web-resources
AN	Access Network	An access network is that part of a communications network which connects subscribers to their immediate service provider.	
AODV	Ad-hoc On-demand Distance Vector	The AODV routing algorithm is for routing data across Wireless Mesh Networks. It is capable of both unicast and multicast routing. It is a reactive routing protocol, meaning that it establishes a route to a destination only on demand. It is defined in IETF RFC 3561.	http://www.ietf.org, http://tools.ietf.org /html/rfc3561
AP	Access Point	A point where users access the system/network, e.g. a base station in a wireless network.	
ARAN	Authenticated Routing for Ad-hoc Networks	A secure routing protocol, ARAN detects and protects against malicious actions by third parties and peers. ARAN introduces authentication, message integrity, and non-repudiation to routing in an ad hoc environment as part of a minimal security policy.	
AsiaCrypt		AsiaCrypt is an IACR conference. The topic is cryptography and cryptographic protocols. IACR also holds the EuroCrypt and Crypto conferences.	http://www.iacr.org/
AuC	Authentication Centre	The AuC is the authentication centre in 2G and 3G cellular networks. The AuC is co-located with a HLR.	
AV	Authentication Vector	The AV is the security credential basis for one challenge-response run in 3GPP and 3GPP2 systems. 3GPP TS 33.102 defines the AV as the following: AV := RAND || XRES || CK || IK || AUTN. 'II' is a symbol for bitstring concatenation.	http://www.3gpp.org /ftp/Specs/html-info /33102.htm
Bluetooth		A short-range wireless specification that allows radio connection between devices within a 10-metre range of each other. Bluetooth is designed as a Personal Area Network (PAN) technology with a wide variety of theoretical uses. Bluetooth is a short-range radio standard and communications protocol primarily designed for low power consumption. Bluetooth, which is a replacement technology for IrDA, provides a unified way to connect devices such as mobile phones, laptops, PCs, printers, digital cameras etc. Bluetooth was named after king Harald Bluetooth, King of Denmark and Norway (born in 910 and died in 985 or 986). The Bluetooth logo is a combination of the Nordic runes Berkanan and Haglaz forming a combined letter/symbol (a bind rune).	https://www.bluetooth.org/
BTS	Base Transceiver Station	The radio base station of a GSM network. It consists of one or more transmitter-receiver unit, each serving one carrier frequency.	http://www.etsi.org
CDMA 2000	Code Division Multiple Access 2000	A family of third-generation (3G) mobile telecommunications standards that use CDMA, a multiple access scheme for digital radio, to send voice, data, and signalling data (such as a dialled telephone number) between mobile phones and cell sites. It is the second generation of CDMA digital cellular. The CDMA2000 standards CDMA2000 1x, CDMA2000 1xEV-DO, and CDMA2000 1xEV-DV are approved radio interfaces for the ITU's IMT-2000 standard and a direct successor to 2G CDMA, IS-95 (cdmaOne). CDMA2000 is standardized by 3GPP2. CDMA2000 is a registered trademark of the Telecommunications Industry Association (TIA-USA) in the United States, not a generic term like CDMA.	http://www.3gpp2.org
Challenge-Response Protocol		Challenge-Response protocols are entity authentication protocols. A principal entity Alice challenges the corresponding principal entity Bob. In order for Bob to respond correctly Bob must compute a reply using a cryptographic function and a personal, secret security credential. There are several distinct types of Challenge-Response protocols, depending on factors such as the type of cryptographic transform used, type of security credential used etc. Challenge-Response protocols can also be classified as unidirectional or mutual (they almost always take place between two principal entities).	
CN	Core Network	Term used for core network nodes in cellular systems. CN nodes include HLR/AuC, VLR/MSC, VLR/SGSN, SMSC, EIR and GGSN.	
COMP 128		An infamous authentication and key agreement algorithm. The original COMP128 is an example implementation for the GSM A3 and A8 cryptographic functions. The COMP128 algorithm is fundamentally flawed and has been known to be so for more than a decade. The algorithm, which is an operator-specific algorithm (contained in the SIM card), is completely unsuitable for its designated taks, yet it is still in use in several GSM/GPRS networks today.	
Cookie (HTTP cookie)		HTTP cookies (or just cookies) are small text objects sent by a server to a web browser. The cookie is returned to the server by the browser upon subsequent visits to the server site. The cookies are used for authenticating, tracking, and maintaining state information about user activities etc. The state information may include user identity, time of last visit, site preferences, the contents of electronic shopping carts etc. The cookies are stored on the client computer and may be a privacy liability. The lifetime of cookies can be set, but the expired cookies may remain on your computer.	
CS	Circuit Switched	A network that establishes a circuit (or channel) between nodes before they may communicate. This circuit is dedicated and cannot be used for other means until the circuit is cancelled/closed and a new one created. If no actual communication is taking place in this circuit then the channel remains idle.	

Acronym /Term	Definition	Explanation	Web-resources
Data Confidentiality (Confidentiality)		The property that information is not made available or disclosed to unauthorised individuals, entities or processes. This property is very closely related to provision of Data Privacy. Defined in 3GPP TS 33.102.	http://www.3gpp.org /ftp/Specs/html-info /33102.htm
Data integrity		The property that data have not been altered in an unauthorised manner. Note that this is a security definition. It differs from the communications definition of data integrity in that the security defintion captures the possibility of malicious intent. Defined in 3GPP TS 33.102.	http://www.3gpp.org /ftp/Specs/html-info /33102.htm
Data Mining		Data mining is the process searching large volumes of data for patterns using various tools to categorize and correlate data into usable information. Data mining can also be defined as "the nontrivial extraction of implicit, previously unknown, and potentially useful information from data".	
Data obfuscation		In security terminology obfuscation is used in the context of concealing the meaning of information or communication by making it more confusing and harder to interpret. Data obfuscation is achieved by applying a transformation function to the data. Data obfuscation transforms may or may not be bijective functions.	
Data origin authentication		Enables the recipient to verify that messages have not been tampered with in transit (data integrity) and that they originate from the expected sender (authenticity). Defined in 3GPP TS 33.102.	http://www.3gpp.org /ftp/Specs/html-info /33102.htm
DDoS	Distributed Denial-of-Service	A distributed and coordinated DoS attack. Usually executed against internet homepages with thousands of (hijacked) computers involved.	
DH	Diffie-Hellman	The Diffie-Hellman key exchange is a cryptographic protocol that allows two parties to establish a shared secret key over an insecure communications channel. The basic DH exchange is unauthenticated and is thus susceptible to MitM attacks.	
DHS	Department of Homeland Security	A US federal state department responsible for coordinating all aspects of homeland security. DHS was created subsequent to the 9/11 terrorist attack.	http://www.dhs.gov /index.shtm
DIAMETER		An AAA (authentication, authorization and accounting) protocol for applications such as network access or IP mobility. It is intended as the successor of RADIUS. The basic concept is to provide a base protocol that can be extended in order to provide AAA services to new access technologies. Diameter is intended to work in both local and roaming AAA situations. Defined in IETF RFC 3588.	http://www.ietf.org, http://tools.ietf.org /html/rfc3588
DoS	Denial-of-Service	A denial-of-service attack (DoS attack) is an attack targeted at the availability of some resource. The attack usually tries to exhaust the capacity of the target in one way or another. Examples include attacks against internet infrastructures like DNS servers, but the most common example would be attacks against high-profile (corporate) homepages. Defined in IETF RFC 4732.	http://www.ietf.org, http://tools.ietf.org /html/rfc4732
DRD	Data Retention Directive	Directive 2006/24/EC of the European Parliament and of the Council of 15 March 2006.	http://europa.eu.int /eur-lex/lex/LexUriServ /site/en/oj/2006/l_105 /l_10520060413 en00540063.pdf
DRM	Digital Rights Management	Any of several technologies used by publishers (or copyright owners) to control access to and usage of digital data (such as software, music, movies) and hardware, handling usage restrictions associated with a specific instance of a digital work.	
DTA	Data Transmission Algebra		
DYI	Dolev-Yao Intruder	The Dolev-Yao Intruder is an exceptionally capable Intruder. It can (and by definition will) capture all messages ever exchanged over any (entity external) interface. It can delete, insert and modify any message at will. Only appropriate and correct use of cryptographic protection can stop the DYI. The DYI will not corrupt the principal entities, but may try to masqeurade as a principal entity.	
EAP	Extensible Authentication Protocol	An authentication framework that enables clients to authenticate with a central server. EAP can be used with several authentication mechanisms (EAP methods), such as EAP-AKA, EAP-SIM, EAP-MD-5, etc. Defined in IETF RFC 3748.	http://tools.ietf.org /html/rfc3748
Eavesdropping		The act of listening in on a conversation (or communication). Eavesdropping is a threat to the privacy of the conversation. Eavesdropping can be prevented in various ways, including the use of the security service Data Confidentiality.	
ECC	Elliptic Curve Cryptography	ECC is a type of public-key cryptography based on the algebraic structure of elliptic curves over finite fields.	

Acronym /Term	Definition	Explanation	Web-resources
EEA	European Economic Area	The agreement creating the European Economic Area (EEA Agreement) was negotiated between the Community, the then Member States, and seven member countries of EFTA. It was signed in May 1992 and came into force 1 January 1994. It was designed to allow EFTA countries to participate in the European Single Market without having to join the EU. The current members (contracting parties) are three of the four EFTA states – Iceland, Lichtenstein and Norway (without Switzerland) – the European Union and the 25 EU Member States.	http://ec.europa.eu /external_relations /eea/index.htm
Entity Authen-tication (Authen-tication)		The provision of assurance of the claimed identity of an entity. Defined in 3GPP TS 33.102.	http://www.3gpp.org /ftp/Specs/html-info /33102.htm
EPC	Electronic Product Code	EPC is a standard for how to tag products electronically.	http://www.epcglobalinc .org/home
ESP	Encapsulating Security Payload	A part of the IPsec framework for Internet security. The ESP extension header provides origin authenticity, integrity, and confidentiality of a packet. It is the preferred IPsec protocol and can provide all the security services IPsec provides. Defined in IETF RFC 4303.	http://www.ietf.org, http://tools.ietf.org /html/rfc4303
ETSI	European Tele-communication Standards Institute	A non-profit membership organisation founded in 1988. The aim is to produce telecommunications standards to be used throughout Europe. The efforts are coordinated with ITU. Membership is open to any European organisation proving an interest in promoting European standards. It was e.g. responsible for the making of the GSM standard. The headquarters are situated in Sophia Antipolis, France.	http://www.etsi.org
GGSN	Gateway GPRS Support Node	Interface between the GPRS wireless data network and other networks such as the Internet or private networks. It supports the edge routing function of the GPRS network. To external packet data networks the GGSN performs the task of an IP router. Firewall and filtering functionality, to protect the integrity of the GPRS core network, are also associated with the GGSN along with a billing function.	http://www.etsi.org, http://www.3gpp.org
GPRS	General Packet Radio Service	An enhancement to the GSM mobile communication system that supports data packets. GPRS enables continuous flows of IP data packets over the system for such applications as web browsing and file transfer. Supports up to 160 kb/s gross transfer rate. Practical rates are from 12 to 48 kb/s.	http://www.etsi.org, http://www.3gpp.org
GPS	Global Positioning System	The Global Positioning System (GPS) is a worldwide radio-navigation system formed from a constellation of 24 satellites and their ground stations. GPS uses these 'man-made stars' as reference points to calculate positions accurate to a matter of metres.	http://www.gps.gov/, http://www.navcen.uscg .gov/gps/default.htm
GSM	Global System for Mobile communi-cations	A digital cellular phone technology system that is the predominant system in Europe, but is also used around the world. Development started in 1982 by CEPT and was transferred to the new organisation ETSI in 1988. Originally, the acronym was the group in charge, Group Special Mobile, but later the group changed name to SMG. GSM was first deployed in seven countries in Europe in 1992. It operates in the 900 MHz and 1.8 GHz band in Europe and 1.9 GHz band in North America. GSM defines the entire cellular system, from the air interface to the network nodes and protocols. As of October 2006, there were more than 2.1 billion GSM users in more than 200 countries worldwide. The ubiquity of the GSM standard makes international roaming very common between mobile phone operators and enables phone users to access their services in many other parts of the world as well as their own country. GSM differs significantly from its predecessors in that both signalling and speech channels are digital, which means that it is seen as a second generation (2G) mobile phone system. This fact has also meant that data communication was built into the system from very early on. GSM is an open standard which is currently developed by the 3GPP.	http://www.gsmworld .com/, http://www.etsi.org, http://www.3gpp.org
Hash / Hash function		A hash function takes an arbitrary length string (the message) and computes a fixed length output string. The output is called the hash, the digest or the checksum. Cryptographic hash functions are different from ordinary hash functions. In the context of this edition of *Telektronikk* we shall only refer to cryptographic hash functions.	http://en.wikipedia.org /wiki/Cryptographic _hash_function
HE	Home Entity	Non-3GPP acronym. Home server. Roughly equivalent to the 3GPP HSS.	
HE	Home Environment	Home Environment: responsible for overall provision and control of the Personal Service Environment of its subscribers.	
HHD	Hand Held Device	A generic name for a pocket-sized computing device, typically utilising a small visual display screen for user output and a miniaturised keyboard for user input (for example, PDA, smartphones etc.).	
HI	Handover Interface		

Acronym /Term	Definition	Explanation	Web-resources
HLR	Home Location Register	The Home Location Register or HLR is a central database that contains details of each mobile phone subscriber that is authorized to use the GSM core network. More precisely, the HLR stores details of every SIM card issued by the mobile phone operator. Each SIM has a unique identifier called an IMSI which is one of the primary keys to each HLR record. The next important items of data associated with the SIM are the telephone numbers used to make and receive calls to the mobile phone, known as MSISDNs. The main MSISDN is the number used for making and receiving voice calls and SMS, but it is possible for a SIM to have other secondary MSISDNs associated with it for fax and data calls. Each MSISDN is also a primary key to the HLR record.	http://www.etsi.org
Homo-morphic crypto-system		Homomorphic crypto-systems have the property that one specific algebraic operation on the plaintext is equivalent to another (possibly different) algebraic operation on the ciphertext. As an example, one can imagine a crypto-system in which addition on ciphertext element is equivalent to multiplication on plaintext elements. This would permit a user, which does not have access to the plaintext, to perform multiplication on the plaintext by executing multiplication operations on the ciphertext. This property is useful in developing Secure Multi-party Computation protocols.	
HSS	Home Subscriber Server	The home subscriber server contains all operative subscriber data, including information on subscribed services, location/roaming information and security credentials. Includes HLR/AuC and AAA services.	http://www.3gpp.org
HUB		A common connection point for devices in a network.	
IACR	International Association for Cryptologic Research	IACR is a non-profit scientific organisation whose purpose it is to further research in cryptology and related fields.	http://www.iacr.org/
Identity Privacy		A privacy service that ensures that the permanent identity of the principal entity is only disclosed to authorized entities. For the case when the entity has multiple identities the service must extend to cover all but anonymous/transient identities.	
Identity Theft		Identity theft can be divided into four categories: A) Financial Identity Theft (using another's name and social security number (or similar) to obtain goods and services, B) Criminal Identity Theft (posing as another when apprehended for a crime), C) Identity Cloning (using another's information to assume his or her identity in daily life), and D) Business/Commercial Identity Theft (using another's business name to obtain credit). An excellent source on identity theft is the Identity Theft Resource Center (http://www.idtheftcenter.org/).	http://www.idtheftcenter .org/
IDS	Intrusion Detection System	A software/hardware tool used to detect unauthorised access to a computer system or network. This may take the form of attacks by skilled malicious hackers, or Script kiddies using automated tools. An IDS is required to detect all types of malicious network traffic and computer usage. This includes network attacks against vulnerable services, data driven attacks on applications, host based attacks such as privilege escalation, unauthorised logins and access to sensitive files, and malware (viruses, Trojan horses, and worms).	
IEEE	The Institute of Electrical and Electronics Engineers	USA based organisation open to engineers and researchers in the fields of electricity, electronics, computer science and telecommunications. Established in 1884. The aim is to promote research through journals and conferences and to produce standards in telecommunications and computer science. IEEE has produced more than 900 active standards and has more than 700 standards under development. Divided into different branches, or 'Societies'. Has daughter organisations, or 'chapters' in more than 175 countries worldwide. Headquarters in Piscataway, New Jersey, USA.	http://www.ieee.org
IEEE 802.11	The IEEE 802 LAN/MAN Standards Committee Working Group for WLAN	Refers to a family of specifications developed by the IEEE for wireless local area networks. It also refers to the Wireless LAN Working Group of the IEEE 802 project. 802.11 specifies an over-the-air interface between a wireless client and a base station or between two wireless clients. The IEEE accepted the specification in 1997. There are several specifications in the 802.11 family, including i) 802.11 – provides 1 or 2 Mbit/s transmission in the 2.4 GHz band; ii) 802.11a – an extension that provides up to 54 Mbit/s in the 5 GHz band. It uses an orthogonal frequency division multiplexing encoding scheme rather than FHSS or DSSS, iii) 802.11b provides 11 Mbit/s transmission in the 2.4 GHz band and was ratified in 1999 allowing wireless functionality comparable to Ethernet; iv) 802.11g provides 20+ Mbit/s in the 2.4 GHz band; v) 802.11z is a method for transporting an authentication protocol between the client and access point, and the Transport Layer Security (TLS) protocol. More variants are also under preparation, including support of 100 Mbit/s traffic flows.	http://www.ieee802.org/11
IEEE 802.11i		IEEE 802.11i is an amendment to the 802.11 standard specifying security mechanisms for wireless networks. Its architecture contains the following components: 802.1X for authentication (entailing the use of EAP and an authentication server), RSN for keeping track of associations, and AES-based CCMP to provide confidentiality, integrity and origin authentication.	http://www.ieee802.org/11

Acronym /Term	Definition	Explanation	Web-resources
IEEE 802.16	The IEEE 802 LAN/MAN Standards Committee Working Group on Broadband Wireless Access Standards	A specification for fixed broadband wireless metropolitan access networks (MANs) that uses a point-to-multipoint architecture. Published on 8 April 2002, the standard defines the use of bandwidth between the licensed 10 GHz and 66 GHz and between the 2 GHz and 11 GHz (licensed and unlicensed) frequency ranges and defines a MAC layer that supports multiple physical layer specifications customized for the frequency band of use and their associated regulations. 802.16 supports very high bit rates in both uploading to and downloading from a base station up to a distance of 30 miles to handle such services as VoIP, IP connectivity and TDM voice and data.	http://www.ieee802.org/16/, http://www.wimaxforum.org/
IEEE 802.1X	IEEE Standards for Local and metropolitan area net-networks – Port-Based Network Access Control	An IEEE standard for port-based Network Access Control; it is part of the IEEE 802 (802.1) group of protocols. It provides authentication to devices attached to a LAN port, establishing a point-to-point connection or preventing access from that port if authentication fails. It is used for certain closed wireless access points, and is based on the EAP, Extensible Authentication Protocol.	http://www.ieee802.org/1/pages/802.1x.html
IETF	Internet Engineering Task Force	A large open international community of network designers, operators, vendors, and researchers concerned with the evolution of the Internet architecture and the smooth operation of the Internet. It is open to any interested individual. The technical work of the IETF is done in its working groups, which are organised by topic into several areas (e.g. routing, transport, security, etc.). Much of the work is handled via mailing lists. The IETF holds meetings three times per year. The IETF working groups are grouped into areas and managed by Area Directors (AD). The ADs are members of the Internet Engineering Steering Group (IESG). Providing architectural oversight is the Internet Architecture Board (IAB). The IAB also adjudicates appeals when someone complains that the IESG has failed. The IAB and IESG are chartered by the Internet Society (ISOC) for these purposes. The General Area Director also serves as the chair of the IESG and of the IETF, and is an ex-officio member of the IAB. The Internet Assigned Numbers Authority (IANA) is the central coordinator for the assignment of unique parameter values for Internet protocols. The IANA is chartered by the Internet Society (ISOC) to act as the clearinghouse to assign and coordinate the use of numerous Internet protocol parameters. IETF's mission statement is given in IETF RFC 3935.	http://www.ietf.org, http://tools.ietf.org/html/rfc3935
IH	Information Hiding	Information hiding addresses two areas of concern: privacy of information from surveillance (steganography) and protection of intellectual property (digital watermarking).	
IIF	Internal Interception Function		
IKE/IKEv2	Internet Key Exchange	IKE is the key exchange protocol for IPsec. It performs entity authentication and key exchange. Defined in IETF RFC 4306.	http://tools.ietf.org/html/rfc4306
IMS	IP Multimedia Subsystem	A standardised Next Generation Networking (NGN) architecture for telecom operators that want to provide mobile and fixed multimedia services. It uses a Voice-over-IP (VoIP) implementation based on a 3GPP standardised implementation of SIP, and runs over the standard Internet Protocol (IP). Existing phone systems (both packet-switched and circuit-switched) are supported. IMS was originally defined by an industry forum called 3G.IP (www.3gip.org) formed in 1999. 3G.IP developed the initial IMS architecture, which was brought to 3GPP for industry standardisation as part of their standardisation work for 3G mobile phone systems in UMTS networks. It first appeared in release 5 (evolution from 2G to 3G networks), when SIP-based multimedia was added. Support for the older GSM and GPRS networks was also provided. 'Early IMS' was defined to allow for IMS implementations that do not yet support all 'Full IMS' requirements. 3GPP2 (a different organisation) based their CDMA2000 Multimedia Domain (MMD) on 3GPP IMS, adding support for CDMA2000.	http://www.3gpp.org, http://www.ietf.org
IMSI	International Mobile Subscriber Identity	The principal subscriber identity in 2G/3G systems. Structure and defintion of IMSI is given both in ITU-T recommendations (E.212) and in 3GPP specifications (TS 23.003). Note that in ITU-T E.212 the acronym is defined as 'International Mobile Station Identity', but the structure is otherwise identical.	http://www.itu.int, http://www.3gpp.org/ftp/Specs/html-info/23003.htm
INI	Internal Network Interface		
Internet		From the commissioning of ARPANET by the US DoD in 1969 the packet switched Internet has gained acceptance and users all over the world. The release of WWW at the end of the 1990s and the browsing possibilities (see WWW) increased the demand for Internet. The interconnection of heterogeneous sub networks of different bandwidths, the best-effort service model and the global end-to-end logical addressing of the internet protocol (IP) has arranged for Internet to be the common information network multiplexing text, pictures, and video as well as packet switched telephony.	
Intruder		In the security literature the term Intruder is reserved for a hostile malicious entity that will intently try to break one or more of the goals of a cryptographic protocol, a protected environment or similar. Sometimes the terms Adversary or Attacker is used for the same purpose.	

Acronym /Term	Definition	Explanation	Web-resources
IP	Internet Protocol	A protocol for communication between computers, used as a standard for transmitting data over networks and as the basis for standard Internet protocols. Originally defined in IETF RFC 791.	http://www.ietf.org, http://tools.ietf.org /html/rfc791
IPsec	IP Security	The IP security architecture consist of a base architetcure and associated security protocols. This includes the ESP and AH security protocols as well as the IKE/IKEv2 key exchange protocols. IPsec is now in its third main revision. Defined in IETF RFC 4301.	http://www.ietf.org, http://tools.ietf.org /html/rfc4301
IR	Infrared	In our context: A technology for short-range data transfer based on optical (infrared) communication. Used in laptops, mobile phones etc. See IrDA.	
IrDA	Infrared Data Association	IrDA is a nonprofit organization whose goal it is to develop globally adopted specifications for infrared wireless communication.	http://www.irda.org/
ISDN	Integrated Services Digital Network	A digital telecommunications network that provides end-to-end digital connectivity to support a wide range of services, including voice and non-voice services, to which users have access by a limited set of standard multi-purpose user-network interfaces. The user is offered one or more 64 kb/s channels.	http://www.itu.int
ISO	International Standardisation Organisation	ISO is a worldwide federation of national standards bodies from more than 140 countries, one from each country. ISO is a non-governmental organisation established in 1947. The mission of ISO is to promote the development of standardization and related activities in the world with a view to facilitating the international exchange of goods and services, and to developing cooperation in the spheres of intellectual, scientific, technological and economic activity.	http://www.iso.org
ITU	International Telecommuni-cation Union	On 17 May 1865, the first International Telegraph Convention was signed in Paris by the 20 founding members, and the International Telegraph Union (ITU) was established to facilitate subsequent amendments to this initial agreement. It changed name to the International Telecommunications Union in 1934. From 1948 a UN body with approx. 200 member countries. It is the top forum for discussion and management of technical and administrative aspects of international telecommunications.	http://www.itu.int
ITU-T	International Telecommuni-cation Union – Standardization Sector	A sector of the ITU whose mission it is to ensure an efficient and on-time production of standards (Recommendations) covering all fields of telecommunications. It was created on 1 March 1993, replacing the former International Telegraph and Telephone Consultative Committee (CCITT).	http://www.itu.int/ITU-T/
k-anonymity		Data records adhere to k-anonymity if each released record has at least (k-1) other records in the release whose values are indistinct over those fields that appear in external data. k-anonymity provides privacy protection by guaranteeing that each released record will relate to at least k individuals even if the records are directly linked to external information.	
KDC	Key Distribution Center	The combination of Authentication Server and Ticket Granting Server of the Kerberos authentication protocol. It is defined in IETF RFC 4120.	http://www.ietf.org, http://tools.ietf.org /html/rfc4120
Kerberos		Kerberos is a computer network authentication protocol which allows individuals communicating over an insecure network to prove their identity to one another in a secure manner. It is designed to provide strong authentication for client/server appli-cations by using secret-key cryptography. Kerberos prevents eavesdropping or replay attacks, and ensures the integrity of the data. Its designers aimed primarily at a client-server model, and it provides mutual authentication – both the user and the server verify each other's identity. Kerberos builds on symmetric key cryptography and requires a trusted third party. It was developed by The Massuchussets Institute of Technology (MIT) in the 1980s and is now maintained by IETF. It is defined in IETF RFC 4120.	http://www.ietf.org, http://tools.ietf.org /html/rfc4120
LAN	Local Area Network	A network shared by communicating devices, usually in a small geographic area. A system that links together electronic office equipment, such as computers and word processors, and forms a network within an office or building.	
LBS	Location Based Service	LBS are services offered to subscribers based on their current location.	
LCS	Location Services		
LEA	Law Enforcement Agency	A Lawful Interception (LI) entity.	
LEMF	Law Enforcement Monitoring Facility	A Lawful Interception (LI) function.	
LI	Lawful Interception	Lawful interception plays a crucial role in helping law enforcement agencies combat criminal activity. Lawful interception of public telecommunications systems in each country is based on national legislation in that country.	http://portal.etsi.org /li/Summary.asp

Acronym /Term	Definition	Explanation	Web-resources
Location Privacy		A privacy service that ensures that the location of the principal entity is only disclosed to authorised entities.	
LR	Local Registers		
MAC	Medium Access Control	The lower of the two sub layers of the Data Link Layer. In general terms, MAC handles access to a shared medium, and can be found within many different technologies. For example, MAC methodologies are employed within Ethernet, GPRS, and UMTS.	
MAC	Message Authentication Code	A MAC function computes a cryptographic signed integrity checksum over an arbitrary length input string under the control of a secret key. MAC functions are quite similar to hash functions, but the MAC function output can only be computed with knowledge of the secret key. MAC functions can be used to provide the message origin authentication and data integrity security services.	http://en.wikipedia.org /wiki/Message _authentication_code
MAP	Mobile Application Part	A protocol that enables real time communication between nodes in a mobile cellular network. A typical usage of the MAP protocol would be for the transfer of location information from the VLR (Visitor Location Register) to the HLR (Home Location Register). Defined in 3GPP TS 09.02 for GSM and in 3GPP TS 29.002 for UMTS.	http://www.3gpp.org/ftp /Specs/html-info /0902.htm, http://www.3gpp.org/ftp /Specs/html-info /29002.htm
MitM	Man-in-the-Middle	In security literature MitM attacks is a class of attacks where the Intruder is located between the legitimate entities. All communication passes through the Intruder, which may selectively delete, deflect, modify and insert messages.	
MIX		The MIX concept is often associated with onion routers, but a MIX can be local and need not route messages. The functionality of a MIX is to disassociate message addresses and message content while still being able to deliver the message to the intended recipient.	
MS	Mobile Station	An MS is the mobile phone. It corresponds to the UE (User equipment).	
MSC	Mobile services Switching Centre	The Mobile services Switching Centre or MSC is a sophisticated telephone exchange which provides circuit-switched calling, mobility management and GSM services to the mobile phones roaming within the area that it serves. This means voice, data and fax services, as well as SMS and call divert. It is located in the core network of a visited network and has an interface towards the radio access network. A Gateway MSC (GMSC) is the MSC that determines which visited MSC the subscriber who is being called is currently located. It also interfaces with the Public Switched Telephone Network. All mobile to mobile calls and PSTN to mobile calls are routed through a GMSC. The term is only valid in the context of one call since any MSC may provide both the gateway function and the Visited MSC function; however, some manufacturers design dedicated high capacity MSCs which do not have any BSCs connected to them. These MSCs will then be the GMSC for many of the calls they handle.	http://www.etsi.org
MSISDN	Mobile Station Integrated Services Digital Network	MSISDN refers to the 15-digit number that is used to refer to a particular mobile station. It is the mobile equivalent of ISDN. The ITU-T recommendation E.164 defines the international numbering plan that MSISDN is based on.	http://www.itu.int
NFC	Near Field Communication Technology	NFC, jointly developed by Sony and Philips, was approved as an ISO/IEC standard on 8 Dec 2003. It was approved as an ECMA standard earlier on. On 18 March 2004 Nokia, Sony and Philips formed NFC-forum to advance NFC development. NFC is essentially about data sharing between devices using short-range radio technologies. NFC holds the promise of bringing true mobility to consumer electronics in an intuitive and psychologically comfortable way since the devices can hand-shake only when brought literally into touching distance.	http://www.nfc-forum.org /home
NGN	Next Generation Network	A network concept that aims at providing a framework to encompass the large variety of existing and emerging protocols and services, facilitate a further evolution of these, decouple the evolution from the underlying network infrastructure, and facilitate the interfacing of a plethora of available media. The rationale behind NGN lies founded in paradigm shifts that have been taking place within the technological solutions and the business models in the telecom industry as a whole. The concept is based on IP-technology and is being specified by ITU-T.	www.itu.int
Nonce	Number used once	A nonce is a cryptographic term used for an element that must provide uniqeness and that will only be used once. A nonce is normally either A) a sequence number, B) a time-stamp, or C) a pseudo-random number.	
P2P	Peer To Peer	A computer network that does not rely on dedicated servers for communication but instead mostly uses direct connections between clients (peers). A pure peer-to-peer network does not have the notion of clients or servers, but only equal peer nodes that simultaneously function as both 'clients' and 'servers' to the other nodes in the network.	

Acronym /Term	Definition	Explanation	Web-resources
Personal Privacy		Personal privacy is a surprisingly difficult term to pinpoint. One aspect is a person's ability to keep details of their daily lives and personal affairs out of public view. It should also include a measure of control over personal information collected by others about themselves. The control right should among other things include the right to restrict the usage and to ensure that the information is correct. Privacy is also sometimes related to a right to being anonymous. Privacy can be seen as an aspect of security and is often achieved by means of security techniques and methods.	
Phishing		Phishing is an activity where a fraudster tries to acquire sensitive/private information. They commonly use social engineering techniques. The most sought after information is usernames, passwords and credit card details etc. A well know phishing technique is to masquerade as a trustworthy website. This includes internet/online banks, auction companies like eBay and other trustworthy sites where you may be compelled to leave sensitive data. Phishing is typically carried out using email, and the users are conned to login or otherwise convey information at a website.	
PKI	Public Key Infrastructure	An arrangement which provides for third-party vetting of, and vouching for user identities. It also allows binding of public keys to users. This is usually carried by software at a central location together with other coordinated software at distributed locations. The public keys are typically in certificates. The term is used to mean both the certificate authority and related arrangements as well as, more broadly and somewhat confusingly, to mean use of public key algorithms in electronic communications. The latter sense is erroneous since PKI methods are not required to use public key algorithms.	
POTS	Plain Old Telephone Service	A very general term used to describe an ordinary voice telephone service. See also PSTN.	
prf	pseudo-random function	A function that generates a stream of pseudo-random numbers.	
Privacy-Preserving		Techniques that ensure that the referred to privacy properties are an invariant property through the execution. Normally the privacy-preserving term is reserved for cryptographic algorithms and protocols that can be formally proven to preserve a given privacy characteristic throughout the execution of the algorithm/protocol.	
PS	Packet Switched	Communication switching method in which packets (units of information carriage) are individually routed between nodes over data links which might be shared by many other nodes. Packet switching is used to optimize the use of the bandwidth available in a network, to minimize the transmission latency (i.e. the time it takes for data to pass across the network), and to increase robustness of communication. The concept of packet switching was developed by Paul Baran in the early 1960s, and independently a few years later by Donald Davies, as described below. Leonard Kleinrock conducted early research and published a book in the related field of digital message switching (without the packets) in 1961, and also later played a leading role in building and management of the world's first packet switched network, the ARPANET.	
Pseudonym		A pseudonym is an alias, used by an individual as an alternative to a person's true name. Use of pseudonyms may provide a measure of identity anonymity.	
Pseudo-Random		Pseudo-randomness (in security) is a property that is normally associated with the characteristics non-predictability, uniqueness and non-repeatability. Statistically the pseudo-random number should (almost always) appear to be uniformly distributed.	
PSK	Pre-Shared Key	In communication security, a secret which was previously shared between the two (or more) parties using an external channel. The characteristics of this secret or key are determined by the system which uses it. It can be a password, a passphrase or a hexadecimal string. This secret is used by all systems involved in the cryptographic processes used to secure the traffic between the systems.	
QoS	Quality of Service	The "degree of conformance of the service delivered to a user by a provider, with an agreement between them". The agreement is related to the provision/delivery of this service. Defined by EURESCOM project P806 in 1999 and adopted by ITU-T in recommendation E.860. [E.860].	http://www.itu.int, http://www.eurescom.de
RADIUS	Remote Authentication Dial-In User Service	An authentication and accounting system used by many (W)ISPs. Then logging into a public Internet service you must enter your username and password. This information is passed to a RADIUS service, which checks that the information is correct, and then authorizes access to the WISP. RADIUS is an AAA protocol. It is intended to work in both local and roaming situations. The RADIUS specification is maintained by a working group of the IETF. Defined in IETF RFC 2865.	http://www.ietf.org/, http://tools.ietf.org /html/rfc2865
RAN	Radio Access Network	A part of a mobile telecommunication system. It implements a radio access technology. Conceptually, it sits between the mobile phone and the core network (CN). Examples are GRAN (GSM RAN), GERAN (GSM/EDGE RAN) and UTRAN (UMTS RAN).	
Random		There are several definitions of what random is intended to mean; notably one has A) an information-theoretic definition; B) a defintion for the statistics field; and C) a defintion for security/cryptography. To distinguish from 'true' randomness one often refers to pseudo-random properties in security terminology.	

Acronym /Term	Definition	Explanation	Web-resources
RFC	Request For Comment	An RFC is a formal document from the Internet Engineering Task Force (IETF) that is the result of committee drafting and subsequent review by interested parties. Some RFCs are informational in nature. Of those that are intended to become Internet standards, the final version of the RFC becomes the standard and no further comments or changes are permitted. Changes can occur, however, through subsequent RFCs that supersede or elaborate on all or parts of previous RFCs.	http://www.whatis.com
RFID	Radio Frequency Identification	RFID is an automatic identification method, relying on storing and remotely retrieving data using devices called RFID tags or transponders. An RFID tag is a small object that can be attached to or incorporated into a product, animal, or person. RFID tags contain antennas to enable them to receive and respond to radio-frequency queries from an RFID transceiver. Passive tags require no internal power source, whereas active tags require a power source.	
RRM	Radio Resource Management		
SA	Security Associations		
SAODV	Secure Ad-hoc On-demand Distance Vector routing	SAODV is an extension of the AODV routing protocol that can be used to protect the route discovery mechanism providing security features like integrity, authentication and non-repudiation.	
SGSN	Serving GPRS support node	SGSN is an exchange which performs packet switching functions for mobile stations located in a geographical area designated as the SGSN area. It is located in the core network of the visited network in 2G/3G systems. It has an interface towardsthe radio access network. The SGSN is the PS equivalent of the VLR/MSC for CS connections.	http://www.3gpp.org, http://www.etsi.org
SHM	Structural Health Monitoring	Structural Health Monitoring is an activity where actual data related to civil structures is observed/measured and registered based on high performance sensors, precision signal conditioning units, broad band analogue-to-digital converters, optical or wireless networks, global positioning systems etc.	http://www.ishmii.org/
SIM	Subscriber Identity Module	The SIM is a subscriber identity module for GSM/GPRS subscriptions. In 2G systems the term SIM is used for a dedicated smartcard with subscriber identity information (including security credentials and algorithms). In 3G systems a SIM is an application running on the UICC (smartcard). Although the terms UICC and SIM are often interchanged, UICC refers to the physical card, whereas SIM (in 3G) refers to a single application residing in the UICC that collects GSM/GPRS user subscription information. The corresponding UMTS subscriber application is the USIM (which is alway present on a UICC). The SIM provides secure storing of the key identifying a mobile phone service subscriber but also subscription information, preferences and storage of text messages. The equivalent of a SIM in UMTS is a Universal Subscriber Identity Module (USIM). Defined in 3GPP specification series 31.	http://www.3gpp.org /ftp/Specs/html-info /31-series.htm
SMC	Secure Multi-party Computation	The research in the field of SMC is often considered to be initiated by Andrew C. Yao in 1982. In short, Yao proposed the so-called millionaire problem in which Alice and Bob are two millionaires who want to find out which is the richer. However, they do not want to reveal how much money they have to each other or to other parties. Solutions to this and other SMC problems tend to rely on use of advanced and sophisticated public-key crypto-system primitives.	
SN	Serving Network	The SN consists of one or more access networks (AN) attached to a core network (CN).	http://www.3gpp.org, http://www.etsi.org
Spyware		The term 'spyware' is used for software that collects personal information about users without their informed consent. The spyware often uses stealth techniques to hide its activity or posing as a legitimate application (which makes it a Trojan).	
SS7	Signalling System #7	A set of telephony signalling protocols which are used to set up the vast majority of the world's PSTN telephone calls.	http://www.itu.int
Steganography		Steganography (literally, covered writing) explores methods to hide the existence of hidden messages.	
TC	Technical Committee		
TETRA	TErrestrial Trunked RAdio	TETRA is a digital trunked mobile radio standard developed by the European Telecommunications Standards Institute (ETSI). The purpose of the TETRA standard was to meet the needs of traditional Professional Mobile Radio (PMR) user organisations, which include utilities, public safety (including the police, the fire brigade, the medical emergency services), government, military, border control, etc.	http://www.tetramou .com/

Acronym /Term	Definition	Explanation	Web-resources
TISPAN	Telecommunication and Internet converged Services and Protocols for Advanced Networking	The ETSI core competence centre for fixed networks and for migration from switched circuit networks to packet-based networks with an architecture that can serve in both. TISPAN is responsible for all aspects of standardisation for present and future converged networks including the NGN (Next Generation Network) and including service aspects, architectural aspects, protocol aspects, QoS studies, security related studies, mobility aspects within fixed networks, using existing and emerging technologies. To a large extent this work is centered around adapting the 3GPP IMS architecture to the TISPAN/ NGN environment. TISPAN is structured as a single technical committee, with core competencies, under which there are Working Groups and Project Teams.	http://www.etsi.org, http://portal.etsi.org /tispan
TLS	Transport Layer Security	Transport Layer Security (TLS) is a protocol that ensures privacy between communicating applications and their users on the Internet. When a server and a client communicate, TLS ensures that no third party may eavesdrop or tamper with any message. TLS is the successor to the Secure Sockets Layer (SSL).	http://www.whatis.com
TMSI	Temporary Mobile Subscriber Identity	TMSI is a 4-octet (byte) unstructured temporary subscriber identity used in the GSM/ GPRS/UMTS systems. Subsequent to initial successful location updating and after encryption has commenced the VLR/SGSN may (should) assign a TMSI to the MS. The TMSI is subsequently to be used as replacement for IMSI. The TMSI is assigned in encrypted form and only used in cleartext, and thus there is no externally apparent binding between the IMSI and the TMSI. In effect this provides a (weak) measure of location- and identity privacy for the mobile subscriber. Defined in 3GPP TS 23.003.	http://www.3gpp.org /ftp/Specs/html-info /23003.htm
ToR	The onion Router	ToR is an anonymity network technology.	http://tor.eff.org/
Trojan		A Trojan is a deceptive program that contains or installs a malicious program (malware) while masquerading as a legitimate application. The term is derived from the classical myth of the Trojan Horse.	
TS	Technical Specification		
TST	Tamper-resistant pseudonym tester		
TTP	Trusted Third Party	In cryptography, an entity which facilitates interactions between two parties who both trust the third party; they use this trust to secure their own interactions. TTPs are common in cryptographic protocols, for example, a certificate authority (CA).	
UE	User Equipment	A UE is a mobile phone (terminal unit, radio unit and smartcard (SIM and/or UICC/USIM)).	
UE	User Entity		
UICC		A physically secure device, an IC card (or 'smart card'), that can be inserted and removed from the terminal equipment. It may contain one or more applications. One of the applications may be a USIM. Defined in 3GPP specification series 31.	http://www.3gpp.org /ftp/Specs/html-info /31-series.htm
UMTS	Universal Mobile Telecommunication System	The European member of the IMT 2000 family of 3G wireless standards. UMTS supports data rates of 144 kb/s for vehicular traffic, 384 kb/s for pedestrian traffic and up to 2 Mb/s in support of in-building services. The standardisation work began in 1991 by ETSI but was transferred in 1998 to 3GPP as a corporation between Japanese, Chinese, Korean and American organisations. It is based on the use of WCDMA technology and is currently deployed in many European countries. As of October 2006 there are more than 90 million subscribers worldwide. The first European service opened in 2003. In Japan NTT DoCoMo opened its 'pre-UMTS' service FOMA (Freedom Of Mobile multimedia Access) in 2000. The system operates in the 2.1 GHz band and is capable of carrying multimedia traffic.	http://www.3gpp.org/, http://www.umts-forum .org
URL	Uniform Resource Locater	A subset of Uniform Resource Identifiers (URI) that identify resources via a representation of their primary access mechanism (e.g. their network 'location'), rather than identifying the resource by name or by some other attribute(s) of that resource. Originally defined by IETF in RFC 1738, later merged with RFC 1808 to RFC 2396 on URN.	http://www.ietf.org
USIM	Universal Subscriber Identity Module	An application residing on the UICC used for accessing services provided by mobile networks, which the application is able to register on with the appropriate security. Defined in 3GPP specification series 31.	http://www.3gpp.org /ftp/Specs/html-info /31-series.htm
UTRAN	UMTS Radio Access Network	Part of the 3G standard UMTS. The UTRAN consists of a set of Radio Network Subsystems (RNS) connected to the Core Network through the Iu-Interface. An RNS consists of a Radio Network Controller (RNC) and a number of base stations called Node Bs. They provide the radio interface Uu towards the User Equipment (UE). Specified by 3GPP. An overall description is found in 3GPP TS 25.401.	http://www.3gpp.org /ftp/Specs/html-info /25401.htm
VLR	Visitors Location Register	VLR is a temporary database of the subscribers who have roamed into the particular area which it serves. Each Base Station in the network is served by exactly one VLR, hence a subscriber cannot be present in more than one VLR at a time. The data stored in the VLR has either been received from the HLR, or collected from the MS. In practice, for performance reasons, most vendors integrate the VLR directly to the V-MSC and, where this is not done, the VLR is very tightly linked with the MSC via a proprietary interface.	

Acronym /Term	Definition	Explanation	Web-resources
WEP	Wired Equivalent Privacy	An implementation of RC4. It is part of the IEEE 802.11 standard (ratified in September 1999) and is a scheme used to secure wireless networks (WiFi). WEP was designed to provide comparable confidentiality to a traditional wired network, hence the name. However, the WEP security protocol is completely broken and attacks will now succeed within a couple of minutes. A new standard, IEEE 802.11i, provides improved security feature. See also WPA/WPA2.	www.ieee802.org, http://www.wifialliance.org
Wi-Fi Alliance		A non-profit international association formed in 1999 to certify interoperability of wireless Local Area Network products based on IEEE 802.11 specification. Currently the Wi-Fi Alliance has 207 member companies from around the world, and over 1000 productshave received Wi-Fi® certification since certification began in March 2000.	http://www.wifialliance .org
WiMAX	Worldwide Inter-operability for Microwave Access	A specification for fixed broadband wireless metropolitan access networks (MANs) that use a point-to-multipoint architecture. Based on the IEEE 802.16 WMAN. Published on 8 April 2002, the standard defines the use of bandwidth between the licensed 10 GHz and 66 GHz and between the 2 GHz and 11 GHz (licensed and unlicensed) frequency ranges and defines a MAC layer that supports multiple physical layer specifications customized for the frequency band of use and their associated regulations. 802.16 supports very high bit rates in both uploading to and downloading from a base station up to a distance of 50 km to handle such services as VoIP, IP connectivity and TDM voice and data.	http://www.ieee802.org /16/, http://www.wimaxforum .org/
WLAN	Wireless Local Area Network	This is a generic term covering a multitude of technologies providing local area net-working via a radio link. Examples of WLAN technologies include Wi-Fi (Wireless Fidelity), 802.11b and 802.11a, HiperLAN, Bluetooth and IrDA (Infrared Data Association). A WLAN access point (AP) usually has a range of 20-300 m. A WLAN may consist of several APs and may or may not be connected to Internet.	
WPA	Wi-Fi Protected Access	An improved version of WEP (Wired Equivalent Privacy). It is a system to secure wireless (Wi-Fi) networks, created to patch the security of WEP. As a successor, WPA implements the majority of the IEEE 802.11i standard, and was intended as an inter-mediate measure to take the place of WEP while 802.11i was prepared.	http://www.ieee802.org, http://www.wifialliance .org
WPA2	Wi-Fi Protected Access 2	An extension to WPA that includes the remaining elements of IEEE 802.11i.	http://www.ieee802.org, http://www.wifialliance.org
WSN	Wireless Sensor Networks	A WSN is a wireless network consisting of spatially distributed autonomous devices using sensors to cooperatively monitor physical or environmental conditions, such as temperature, sound, vibration, pressure, motion or pollutants.	

Kaleidoscope

The History Behind the Probability Theory and the Queuing Theory

KJELL STORDAHL

Kjell Stordahl is Senior Advisor in Telenor Nordic Fixed

The author defended his DrPhilos thesis "Long-Term Telecommunications Forecasting" at NTNU, Trondheim ultimo 2006. His chosen lecture was: *"Are the odds against you? The history of how gambling initiated the theory of probability, and how the theory can be used to improve your odds"* (Stordahl, 2006). This article documents part of the lecture and extends the perspectives by drawing lines to establishment of the teletraffic/queuing theory.

Background

Humans have practised gambling at all times. The archaeologists have made excavations in prehistoric sites and found large numbers of roughly dice-shaped bones. Different types of games, sports events, other types of events and gambling are connected because it has always been challenging to make bets on different outcomes of a game.

Playing with dice, relief from ancient Rome

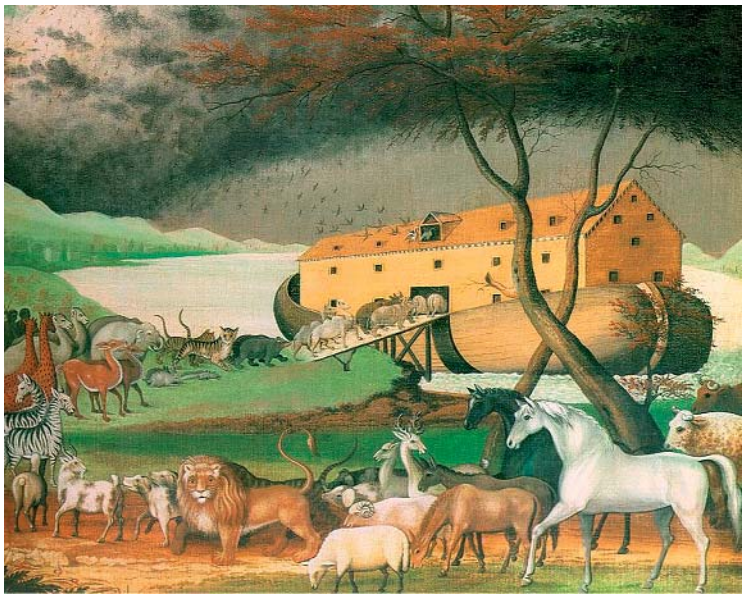
Experiences and simple statistics used more or less unconsciously made in old times the basis for the gamblers and their betting. Until the 16th century the mathematics was not applied on gambling and probability problems.

This paper shows how gambling problems initiated the mathematical theory of probability and gives an overview of the establishment of the mathematical theory of probability. The lines are drawn from the mathematical theory of probability to the establishment of the queuing/teletraffic theory more than 200 years later. And the pioneers in developing the queuing/ teletraffic theory were Agner Krarup Erlang and Tore Olaus Engset!

(Stordahl, 2006) identified that The Gambler's Ruin Problem was solved by using the same difference equations as for the M/M/1 queuing systems, only 200 years earlier. This paper investigates and explains the incitements for the development of the queuing/teletraffic theory which was mainly caused by introduction of telephone switching systems.

Gambling, caricature of an early roulette table, ca 1800

ISSN 0085-7130



Early queuing system. Edvard Hicks (1780 – 1849): "Noah's Ark"

Before starting, a quotation from the famous mathematician Marquis Pierre-Simon de Laplace in his book *Thèorie Analytique des Probabilitiés* (Laplace, 1812) is in its place: *"It is remarkable that a science which began with consideration of games of chance should have become the most important object of human life ... The most important questions of life are, for the most part, really only problems of probability."*

The History Behind the Evolution of the Mathematical Theory of Probability

The most important sources used in this chapter are: (Todhunter, 1865), (Ore, 1953), (Ore, 1960), (King, 1990) and (Mahoney, 1994). The book by Isaac Todhunter from 1865, containing 1062 notes, is considered to be the real bible of the mathematical theory of probability. The book has on different subjects been shaded and supplemented by Anders Hall (Hall, 1990).

The very early start of the history of the mathematical theory of probability is strongly influenced by Gerolamo Cardano, Blaise Pascal, Pierre de Fermat and Christian Huygens.

Gerolamo Cardano (1501 – 1576)

Gerolamo Cardano is said to be the father of the mathematical theory of probability. Øystein Ore has written a very interesting book (Ore, 1953) where he analyses the many abilities of the unusual man and scientist Cardano. The last part of the book contains Cardano's book *Liber de Ludo Aleae (The Book on Games of Chance)* which Ore had translated from Latin.

Gerolamo Cardano's father, Fazio Cardano, was a lawyer, but he had excellent knowledge in medicine and mathematics. He was also consulted by Leonardo da Vinci regarding analysis in geometry.

The life of Gerolamo Cardano is very well described in his autobiography *De Propria Vita*. He was a well educated medical doctor. While studying he gambled to finance his studies. He also carried on gambling for many years. Cardano had many different interests and he published books in mathematics, physics, astronomy, probability theory, moral, upbringing, music, chess, dreams, death; but most of his books dealt with medical questions. In his autobiography he counted 131 printed works after having burned 170 manuscripts which he judged not to be good enough. Cardano's first book *On the Bad Practices of Medicine in Common Use* was published in 1536. The book was written because Cardano applied for but did not get a position at the hospital in Milan. However, after the publication the medical doctors were frightened and gave him a position. During the course of a few years he got the top position at the hospital. He was a skilful medical doctor and was often used by the aristocracy. He was also a unique debater who was impossible to beat in the duels held at that time.

In 1545 Cardano issued *Ars Magna*, a textbook in arithmetic, where the solutions of the third and fourth degree equation were published for the first time. The solution of the third degree equation was originally found by Scipione del Ferro about 1500, but he did not publish his solution. At that time teachers at the universities in Italy could be challenged for their position through competitions where each duellist put up a set of questions to be solved by the other. Hence, it was better to have and to hide knowledge than to

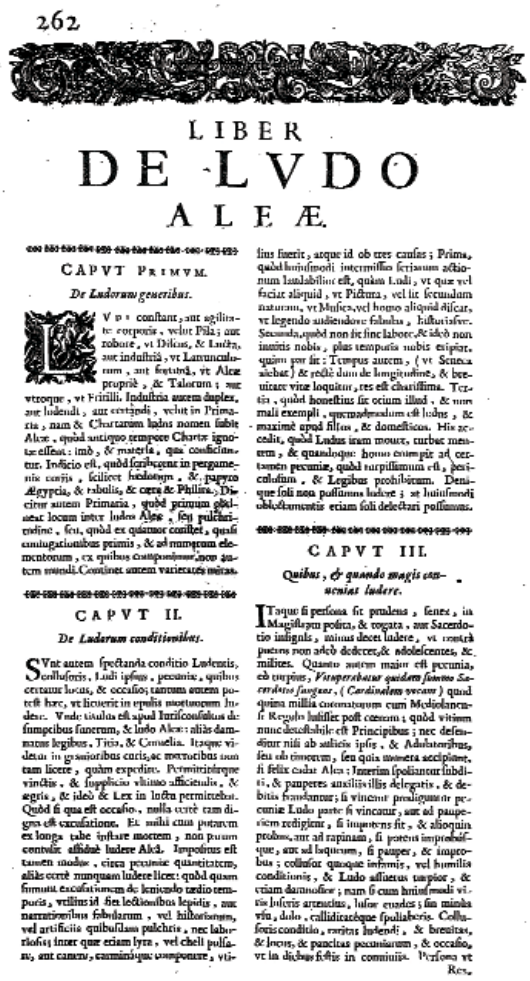
Gerolamo Cardano (1501 – 1576) is said to be the father of the mathematical theory of probability

share it. The solution of the fourth degree equation was found by Cardano's brilliant scholar Lodovico Ferrari.

Cardano wrote several books about gambling. The book *Liber de Ludo Aleae* is a handbook for gamblers. The book shows that Cardano was an experienced gambler. A lot of advice is presented about tricks and cheating, and recommendations are given to avoid cheating. Cardano applies basic principles for the probability theory. He defines probability as the number of favourable (outcomes) divided by the number of possible (outcomes). He states that the probability of a set of independent events is equal to the product of the probability of each of the events. He also touches the mathematical expectation and the law of large numbers. Cardano shows a complete sample space for throws with two and with three dice. His methods would in principle solve the Chevalier de Mère's problem, which, as will be seen was discussed by de Mère and Pascal about 100 years later.

The main part of the manuscript for the book was written at an early stage of Cardano's career, but some parts were included later. Unfortunately, he was not allowed to publish the book because in 1570 he was arrested by the Inquisition and denied further publishing. Regrettably, the book *Liber de Ludo Aleae* was not published until 1663, and then as a minor part of a large ten cover volume of Cardano's publications.

Therefore, *Liber de Ludo Aleae* did not make the impact on the evolution of the mathematical theory of probability as it could have done!



262

LIBER

DE LVDO

ALEÆ.

CAPVT PRIMVM.

De Ludorum generibus.

VBI constant, aut agilitate corporis, velut Pila; aut robore, vt Discus, & Lucta, aut industriâ, vt Lanuncularum, aut fortunâ, vt Aleæ propriè, & Talorum; aut vtroque, vt Fritilli. Industria autem duplex, aut ludendi, aut certandi, velut in Primacia, nam & Chartarum ludus nomen habit Aleæ, quòd antiquo tempore Charta ignotæ essent: imò, & materia, quæ constituitur. Inficio est, quòd scribentes in pergamenis corijs, scilicet hœdorum, & pappyro Ægyptia, & tabulis, & cera & Philira. Dicitur autem Primacia, quòd primum obtineat locum inter ludos Aleæ, seu pulchritudine, seu, quòd ex quatuor consistet, quasi quadrigationibus primis, & ad numerum elementorum, ex quibus compositum, quo mundi. Continet autem varietates octo.

CAPVT II.

De Ludorum conditionibus.

SVnt autem spectanda conditio Ludentis, deinde loci, Ludi ipsius, pecuniæ, quibus certatur locus, & occasio; tantum autem potest hæc, vt liceriat in epulis mortuorum Indices. Vnde vitiata est apud Iurisconsultos assumptibus funerum, & ludo Aleæ: aliàs damnatus legibus, Titio, & Cæsellio. Itaque videtur in grauioribus curis, ac mœrotibus non tam licere, quàm expedire. Permittitúrque vinctis, & supplicio vltimo afficiendis, & ægris, & ideò & Lex in lusta permittebat. Quòd si qua est occasio, nulla certè tam digna est excusatione. Et mihi cum putarem ex longa tabe instare mortem, non parum contulit assidue ludere Aleâ. Impotius est tamen modus, circa pecuniæ quantitatem, aliàs certè nunquam ludere licet: quòd quam summa excusationem de leniendo tædio temporis, vellim id fieri lectionibus lepidis, aut narrationibus fabularum, vel historiarum, vel artificiis quibusdam pulchris, nec laboriosis; inter quæ etiam lyra, vel chelys pulsanda, aut cantus, cantusque componere, vtilius fuerit, atque id ob tres causas; Prima, quòd huiusmodi intermissis seriarum actionum laudabilius est, quàm Ludi, vt quæ vel faciat aliquid, vt Pictura, vel sit secundum naturam, vt Musica, vel homo aliquid discat, vt legendo audiendoue fabulas, historiasve. Secunda, quòd non sit sine labore, & ideò non inutilis nobis, plus temporis nobis eripiat, quàm par sit: Tempus autem, (vt Seneca inquit) & rectè dum de longitudine, & breuitate vitæ loquitur, res est charissima. Tertia, quòd honestius sit otium illud, & non malis exempli, quemadmodum est ludus, & maximè apud filios, & domesticos. Hinc accidit, quòd Ludus iram mouet, turbat mentem, & quandoque homo exemplo ad certamen pecuniæ, quòd turpissimum est, periculosum, & Legibus prohibitum. Denique soli non possumus ludere; at huiusmodi obiectamentis etiam soli delectari possumus.

CAPVT III.

Quibus, & quando magis conueniat ludere.

ITaque si persona sit prudens, senex, in Magistratu posita, & rogata, aut Sacerdotio insignis, minus decet ludere, vt contrà pueros non adeò dedecet, & adolescentes, & milites. Quanto autem maior est pecunia, eò turpius, *Vterque habetur quidam summo Sacerdotio fungens*, (*Cardinalem vocant*) quod quinta millia coronatorum cum Mediolanensi Regulo ludisset post cœnam; quòd vitium nunc detestabile est Principibus; nec defuerunt nisi ab adiectis ipsis, & Adulatoribus, seu ob timorem, seu quia monetæ sectipiæ, si felix cadat Alea: Interim spoliantur libidinis, & pauperes auxiliis illis delegatis, & debitis fraudantur; si vincunt prodigunturque pecuniæ Ludo parte si vincatur, aut ad pauperiem redigitur, si inquietus sit, & alioquin probus, aut ad rapinam, si potens improbissimque, aut ad iniquum, si pauper, & improbus; colluitur quoque infamis, vel humilis conditionis, & Ludo assuetus turpior, & etiam damnosior; nam si cum huiusmodi vitiis lusoris arrectus, lusor eruditus; sin minùs vtu, dolo, calliditateque spoliaberis. Collusoris conditio, varietas ludendi, & breuitas, & locus, & punctus pecuniarum, & occasio, vt in dictum fuit in communi. Personæ vt Res.

Gerolamo Cardano's book Liber de Ludo Aleae *was not published until 1663, then as part of* Opera Omnia*, a ten cover volume of Cardano's publications*

Blaise Pascal (1623 – 1662)

Blaise Pascal (1623 – 1662)

Blaise Pascal was taken very ill when he was one year old. Sickness followed Pascal through his life and he only lived to be 39 years old. When he was 16 he published a remarkable thesis on conic cuts. At the age of 18 he made the world's first calculation machine for addition and subtraction – in fact for helping his father with tax calculations, which was part of his job. He also carried out a series of pressure measurements and concluded on the existence of vacuum. Pascal was also an excellent author with a specific talent for polemics.

The literary style of Pascal was influenced by Antoine Gombaud Chevalier de Mère (1607 – 1684) who he got acquainted with in 1651/1652. Many books on the history of the mathematical theory of probability start with: *"A gambler named Chevalier de Mère presented two gambling problems to Blaise Pascal"*. (Ore, 1960) tells that "Chevalier de Mère would have turned in his grave at such a characterisa-

The dice problem: How many times do you have to throw two dice to have a probability higher than 0.5 to get at least one double 6 in the sequence?

tion of his main occupation in life". Chevalier de Mère had received a good classical education and had experience from the army. He served in the court in Paris. He was a philosopher and a writer and he rapidly became a prominent figure at the court of Louis XIV where he was adviser in delicate situations and arbiter in conflicts.

Chevalier de Mère made Pascal aware of *The dice problems* which had been well known during the last centuries. One of the dice problems is described as follows: *How many times do you have to throw two dice to have probability higher than 0.5 to get at least one double 6 in the sequence?* Pascal solved the problem in the following way:

He stated that the probability not to get a double 6 in one throw is *35/36*. Then he postulated the same as Cardano did a hundred years earlier, that throws with dice are independent events and expressed that the probability not to get one double 6 in n throws is $(35/36)^n$. Hence, the probability to get at least a double 6 in n throws is $p_n = 1 - (35/36)^n$. Then, Pascal calculates $p_{24} = 0.491$ and $p_{25} = 0.506$. Hence, the limit is between throw 24 and throw 25.

At that time Pascal made contact with Fermat to get confirmation of his theories and this process is considered by many to be the start of the evolution of the mathematical theory of probability.

Pierre de Fermat (1601 – 1665), contemporary engraving

Pierre de Fermat (1601 – 1665)

Pierre de Fermat came from a wealthy merchant family on his father's side and a lawyer family on his mother's side. He studied law at the universities of Toulouse and Orléans and mathematics at the university of Bordeaux. He made a career as a lawyer and got continuously higher positions. It could be said that his professional career was as a lawyer, but he had mathematics as a lifelong hobby! Fermat was known for showing his mathematical results, but he did not always show his proofs. The reason was that he did not consider communicating his proofs to be his primary tasks in life. Significant parts of Fermat's scientific work have been found in the margins of manuscripts and in letters to his friends. In other words, he was not very interested in documentation and publication of his mathematical works. However, at a later stage of his life, in August 1654, he suggested to Carcavi that Carcavi and Pascal should publish his scientific work. His son made this possible 14 years after his death, in 1679.

In 2000, the 'World Mathematical Year', The Czech Republic issued a stamp showing Fermat's last theorem, also showing that Andrew Wiles proved it in 1995

Fermat became known for making the foundation of the analytical geometry. He made significant contributions to the calculus through calculations of tangent, maximum and minimum and to the number theory. Fermat is also famous for his *last theorem* (however from 1637) where he states that there exist no integers which satisfy $x^n + y^n = z^n$ where $n > 2$. The proof is said to be made in the margin of one of his manuscripts.

Anyhow, the statement generated a fantastic story of how the best brains in mathematics over a period of more than 350 years tried to develop this proof (Singh, 1997). Then in 1993, based on seven years work in isolation, Professor Andrew Wiles from Princeton University, NJ, USA, publishes the proof at a mathematical conference in Cambridge. The event caused enormous publicity. However, a month later, it was shown that there was a 'hole' in the proof – it was incomplete. It was a catastrophe for Andrew Wiles.

He had isolated himself like the Pythagorean did 2000 years earlier when they developed mathematics as a religion – but only for the initiated! In spite of the enormous pressure from the media, Andrew Wiles succeeded one year later to complete the proof. In the completion he also used the Selmer groups. The Selmer groups were developed at the beginning of the 1950s by the Norwegian professor Ernst Selmer, who regrettably died on 8 November 2006.

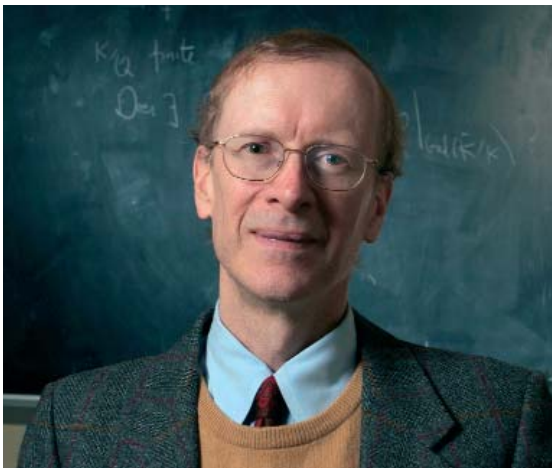
Start of the Mathematical Theory of Probability

The start of the mathematical theory of probability is by many considered to be the exchange of letters between Pascal and Fermat in 1654.

Pascal wanted at that time to get confirmation on some of his proofs on gambling problems. He made contact with the well known mathematician Roberval, but he did not get any support, only criticism. Roberval was described as "the greatest mathematician in Paris, and in conversation the most disagreeable man in the world". Pascal then consulted Fermat, who was living in Toulouse. Fermat was isolated from the mathematical environment and was happy to have contact with Pascal, which was reciprocated. There then followed an exchange of minimum seven letters, and this is by many considered as the real start of the mathematical theory of probability.

In the correspondence Fermat confirms Pascal's solutions on the dice problems.

However, the correspondence starts with the classical *Point problem,* which was well known from several centuries back. So far, nobody had found the solution.



Andrew Wiles, professor at Princeton University, NJ, USA, finally was able to prove Fermat's last theorem in 1995

There were different variants and wrappings of the problem. In the following is shown Pascal's solution to a simple Point problem.

Two players A and B each put 32 gold coins (also called pistols) in the pot. The first player who gets three points has won the game and will get all the gold coins. The winner of each round gets one point. Each player has equal probability, 1:2, to win a round. However, the game is disrupted when player A has two points and player B one point. The question is: How should the pot be divided in a fair way?

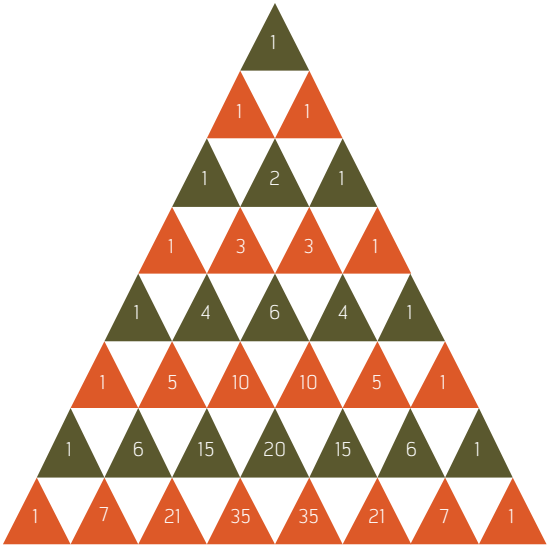
Pascal allocates probabilities to different realisations and reasons as follows: The probability for player A to win the next round is 1:2. Because of that he should have 32 gold coins. The probability for player B to win the next round is 1:2. Then, player A and B both have two points and they should of course divide the remaining 32 gold coins equally, getting 16 gold coins each. Hence, player A should have a total of 48 gold coins and player B 16.

Fermat generalises the point problem and finds solutions for more complicated cases also when there are several players. He uses permutations, which can be used as long as the probability of winning is equal for each player. Fermat also finds some failures in Pascal's reasoning which he corrects.

The point problem. A fair division of gold coins (pistols)

Then, Pascal develops solutions by using Pascal's triangle. Pascal's arithmetic

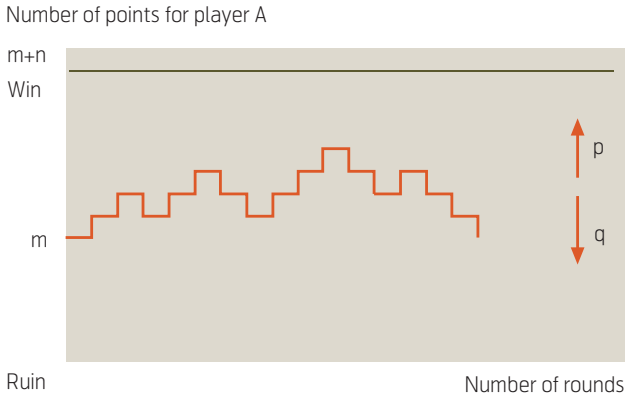
ISSN 0085-7130



Pascal's triangle is a geometric arrangement of the binomial coefficients

triangle was well known, but got Pascal's name because he discovered new properties of the triangle which were previously not known. Pascal shows how binomial coefficients are used to calculate the probability of player A to win (and B) when A needs *m* points and B needs *n* points. Fermat also did the same using the Binomial distribution.

The exchange of letters stopped at the end of 1654 when Pascal sacrificed himself, as the rest of his family, for the Jansenism[1)]. His last years were spent mainly working with religious questions and in 1656 he wrote *Lettres Provinciales* opposing the Jesuits' attack on Jansenism.



The Gambler's Ruin problem

The Gambler's Ruin Problem

The correspondence between Pascal and Fermat was renaissanced a short period in 1656. Here, Pascal puts a question forward to Fermat which was known as the Gambler's Ruin problem. Fermat and Pascal solved the problem for some distinct values of the parameters, but it is not known that they solved the general problem.

The Gambler's Ruin problem is as follows: Player A starts with *m* points and player B with *n* points. The probability of player A winning one round is *p*, while the probability of player B winning a round is $q = 1 - p$. The winner of one round receives one point from the other player. The question is: What is the probability of player A winning the game; that is, to get $n + m$ points? And in general, what is the probability of player A winning the game when he has *i* points?

Christiaan Huygens (1629 – 1695) was next to improve the mathematical theory of probability significantly. In 1657 he published *Libellus De Ratiociniis in Ludo Aleae (The Value of all Chances in Games of Fortune)*, a book on probability theory. The book was published six years before Cardano's book and con-

Christiaan Huygens (1629 – 1695), engraving of Frederik Ottens, 18th century

1) *Jansenism was a branch of Catholic thought that emphasized original sin, human depravity, the necessity of divine grace, and predestination. Originating in the writings of the Flemish theologian Cornelius Otto Jansen, Jansenism formed a distinct movement within the Roman Catholic Church from the 16th to 18th centuries, but was condemned by the Roman Catholic Church as heretical (http://en.wikipedia.org/wiki/Jansenism)*

tains, among other things a precise definition of the concept of mathematical expectation. Huygens also puts up five unsolved probability problems; one of them, Huygens fifth problem, is The Gambler's Ruin problem which Huygens only solved for some values of the parameters.

Some more years passed before the general solution was found. In fact the solution was found by different approaches by James Bernouilli (1708), Montmort (1708), de Moivre (1712) and Struyck (1716), see (Hald, 1990). The process is a random walk with absorbing barriers in 0 and $n + m$.

A complete proof based on difference equations was first given by Struyck (1716), see (Hald, 1990, page 203). He finds the explicit solution of the difference equations first for $n = m$ and then for n different from m. Let $p(i)$ be the probability for player A to win the game given that he has i points. Then, the difference equations can be expressed by:

$$\begin{aligned} p(i) &= p\,p(i-1) + q\,p(i+1), \\ &\quad i = 1, 2, ..., m+n-1 \qquad (1) \\ p(0) &= 0 \\ p(m+n) &= 1 \end{aligned}$$

The first equation expresses that the probability of winning the game given that the player has i point is equal to p multiplied by the probability to win given that the player has i - 1 points pluss q multiplied by the probability that the player has $i + 1$ points. The two last equations are the edge conditions stating that the probability of player A winning is 0 if he has no points left (he is ruined) and the probability of winning when he has $n + m$ points is of course 1.

Now, we know that $p + q = 1$. The equation will not be changed when the left hand side is multiplied with $(p + q)$. Hence the equation is:

$$\begin{aligned} (p+q)\,p(i) &= p\,p(i-1) + q\,p(i+1), \\ &\quad i = 1, 2, ..., m+n-1 \qquad (2) \end{aligned}$$

This difference equation is exactly the same as the difference equation which describes the M/M/1 queuing system in statistical equilibrium. *However, the development of the queuing theory was not in place until 200 years later. The history of the evolution and the incitements for the evolution is treated in the last part of the paper.*

Queuing Models and Queuing Theory

Queuing theory is the mathematical study of waiting lines or queues. The theory enables mathematical analysis of several related processes, including arriving at the queue, waiting in the queue, and being served by the server(s) at the front of the queue. The theory permits the derivation and calculation of several performance measures including the average waiting time in the queue or the system, the expected number waiting or receiving service and the probability of encountering the system in certain states, such as empty, full, having an available server or having to wait a certain time to be served.

Birth-death processes have many applications in demography, queuing theory, and in biology, for example to study the evolution of bacteria. The state, i, of the process represents the current size of the population. The transitions are limited to births and deaths. When a birth occurs, the process goes from state i to $i + 1$. When a death occurs, the process goes from state i to state $i - 1$.

The simplest form of queuing models are based on the birth and death process, where the birth process describes the inter-arrival time (time between two arrivals) to the queue and the death process describes the service or holding time in the queue.

For queuing theory, it has been found convenient, if possible, to work with probability distributions which exhibit the memorylessness property, as this commonly simplifies the mathematics involved.

The memorylessness property is often denoted a Markovian property and a process with a Markovian property is called a Markov process, which means that the probability distribution of future states of the process, given the present state and all past states, depends only upon the present state and not on any past states.

Andrej Markov (1856 – 1922)

As a result, queuing models are frequently modeled as Poisson processes through the use of the exponential distribution.

The Markov process is named after the famous Russian mathematician Andrej Markov (1856 – 1922). Andrej Markov was created honorary doctor at the University of Oslo at the Abel jubilee in 1902 (Nils Henrik Abel (1802 – 1829))

The Poisson process, discovered by the French mathematician *Siméon-Denis Poisson* (1781 – 1840) is a pure-birth process, the simplest example of a birth-death process. The Poisson distribution is a discrete probability distribution. It expresses the probability, $p(i, \lambda)$, of a number of events i, occurring in a fixed period of time, if these events occur with a known average rate λ, and are independent of the time since the last event.

$$p(i,\lambda) = \frac{e^{-\lambda}\lambda^{i}}{i!}$$

Suppose that the inter-arrival time is described by an exponential distribution with parameter λ (traffic intensity), and the holding time is described by an exponential distribution with parameter μ. Then the transient behavior of the queuing system is expressed by:

$$p_i'(t) = \lambda\, p_{i-1}(t) + (\lambda + \mu)p_i(t) + \mu\, p_{i+1}(t) \qquad (3)$$

where $p_i'(t)$ is a derivative to $p_i(t)$ which is the probability to have i in the queue system at time t. The system is described as a function of time and can be solved when we know the starting value at time 0.

Suppose that the system reaches statistical equilibrium. Then the solution is independent of the starting values. In addition, it has a balance between inter-arrivals and services which implies $\lambda/\mu < 1$. Then $p_i'(t) = 0$.

Letting $p_i(t) = p_i$, we get:

$$(\lambda + \mu)p_i = \lambda p_{i-1} + \mu p_{i+1} \qquad (4)$$

which is identical to the Gambler's Ruin problem equation (2), but with different notations.

This queuing system is denoted M/M/1: Exponential inter-arrival time and holding time and one server. The classification of queuing systems follows Kendell's definition (Kendell, 1953). The solution is found by expressing all the $\{p_i\}$ as a function of p_0 and then normalize based on the summing up of all the probabilities to 1. The same procedure is done for the Gambler's Ruin problem, but the edge conditions are also taken into account.

To show the equality in the solutions the following notations are used:

$$p(i) = p_i \qquad (5)$$

$$\rho = \lambda/\mu = p/q \qquad (6)$$

$$K = n + m \qquad (7)$$

Solutions of different queuing systems and the Gambler's Ruin problem:

The Gambler's Ruin problem:

$$p(i) = (1 - \rho^i)/(1 - \rho^K) \qquad (8)$$

M/M/1: $$p(i) = (1 - \rho)\, \rho^i \qquad (9)$$

M/M/1/K: $$p(i) = \rho^i\, (1 - \rho)/(1 - \rho^{K+1}) \qquad (10)$$

M/M/∞: $$p(i) = (\rho^i / i!)\, e^{-\rho} \qquad (11)$$

M/M/K/K:

$$p(i) = (\rho^i/i!)/\left(\sum_{k=1}^{K} \rho^k/k!\right) \qquad (12)$$

Erlang's B loss formula:

$$p(K) = (\rho^K/K!)/\left(\sum_{k=1}^{K} \rho^k/k!\right) \qquad (13)$$

Here, A/B/C/D follows the notation of (Kendell, 1953) and (Kleinrock, 1975) where:

A: Interarrival time distribution
B: Service time distribution
C: Number of servers
D: Waiting room capacity

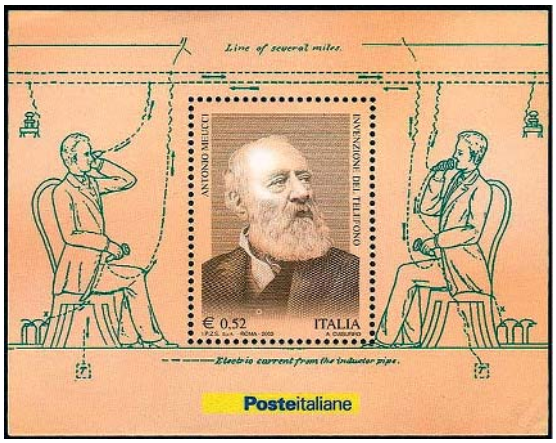
It could be noted that the solution of queuing models with more than one server uses $(i+1)\mu$ instead of μ in equation (4).

It is interesting to note that most books in queuing theory use equation (4) as a standard formula because it is derived from the transient equation (3). The stationary state equations can be interpreted as follows: The traffic stream out of state i is equal to the traffic stream into state i. Looking at the original work of Erlang (Erlang 1917) (Brockmeyer, 1948) he uses another approach. Instead of assuming that the traffic stream out of a state is equal to the traffic stream into the state, he postulates that the traffic stream is equal both ways between a cut of states. He then gets the simplified equation:

$$\lambda\, p(i) = (i + 1)\mu\, p(i + 1) \qquad (14)$$

where he also uses $(i + 1)\mu$ as a more general expression. Arne Jensen uses the same approach in his paper (Jensen, 1954). A complete proof is given in (Morris, 1961). The possibility of using the cut is a much more powerful approach for modelling more complicated stationary queuing systems. This is shown in (Stordahl, 1972).

Let us go back and look at the real incentives for developing the queuing and teletraffic theory. The whole thing started with the telephone!



Antonio Meucci (1808 – 1889), the real inventor of the telephone was portrayed on an Italian stamp in 2005

The Telephone and the Manual Switches

The Italian Antonio Meucci (1808 – 1889), who had already created the first model of a telephone in Italy in 1834, tested electric transmission of the human voice in Cuba in 1849 and demonstrated his electric telephone in New York, USA in 1850. He had paid for a 'caveat' for the telephone in 1871. In the summer of 1872 Meucci asked Edward B. Grant (Vice President of American District Telegraph Co. of New York) permission to test his telephone apparatus on the company's telegraph lines. He gave Grant a description of his prototype and copy of his caveat. Up to 1874 Meucci only had enough money to renew his caveat while looking for funding for a true patent. After waiting two years without receiving an answer, Meucci went to Grant and asked him to be given back his documents, but Grant answered that he had lost them. The same year the caveat expired because Meucci lacked the money to renew it.

> *In 2002 the American Congress announced that Antonio Meucci (not Alexander Graham Bell) was the real inventor of the telephone.*
> (Kunnskapsforlaget, 2007)

Meanwhile, in 1867 the following could be read in an American newspaper: "The 46 years old, Joshua Coppersmith has been arrested in New York for attempting to extort funds from ignorant and superstitious people by exhibiting a device which he says will convey the human voice over metallic wires, so that it will be heard by the listener at the other end. He calls the instrument a *telephone*, which is obviously intended to imitate the word 'telegraph', and win the confidence of those who know of the success of the latter instrument without understanding the principles on which it is based. Well-informed people know that it is impossible to transmit the human voice over wires as may be done with dots and dashes and signals of the Morse Code, and that were it possible to do so, the thing would be of no practical value. The authorities who apprehended this criminal are to be congratulated, and it is to be hoped that it may serve as an example to other conscienceless schemers who enrich themselves at the expense of their fellow creatures."

Eight years later Alexander Graham Bell and his assistant Thomas Watson started to work on a device they called *A musical telegraph*, and on 10 March 1876 they succeeded in completing the device, which was eventually named the telephone. Bell applied for a patent for the invention and got it, but it was a close race since another American, Elisha Gray, applied for patent of a similar device *only two hours later!*

The telephone was demonstrated at the World exhibition in Philadelphia in May 1876. In one of the juries at the exhibition was a Norwegian, Joak Andersen,

Alexander Graham Bell, the later inventor of the telephone (1847-1922). The picture shows a well-known scene were Bell speaks on the phone between New York and Chicago in 1892. (Gilbert H. Grosvenor Collection, Prints and Photographs Division, Library of Congress)

Alexander Graham Bell (1847 – 1922) was a scientist and innovator. Born and bred in Scotland, he emigrated to Canada in 1870, and the following year to the United States. Bell is widely acclaimed for developing and patenting the telephone (at the same time but independently from Elisha Gray, and with prior efforts from Antonio Meucci and Philipp Reis). In addition to Bell's work in telecommunications, he was responsible for important advances in aviation and hydrofoil technology.

ISSN 0085-7130

Lars Magnus Ericsson (1846 – 1926), the founder of L.M. Ericsson

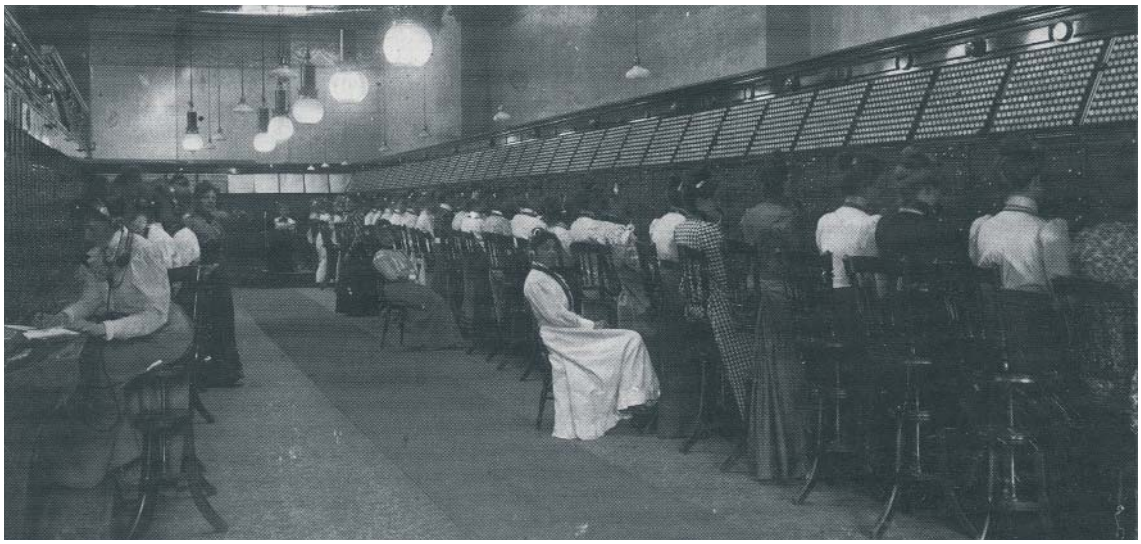
vice-consul to Denmark, who got two telephones which he sent to his son in Ålesund. However, the first public demonstration of the telephone in Norway was done in Bergen on 22 July 1877. The painter Johan Eimrich Rein had received two telephones from a friend who got them from Alexander Graham Bell. Then the engineer Jens Hopstock started a tour of Norway where he demonstrated the new invention. He also demonstrated the telephone in Stockholm and was even invited to King Oscar II to demonstrate it. Jens Hopstock was later appointed The International Bell Telephone Company's representative for Scandinavia (Bestorp, 1990).

Already in the autumn 1877 imitations of the telephone were made by Siemens & Halske in Germany. This prototype was the inspiration to the young instrument maker *Lars Magnus Ericsson* who started production of telephones later that year. He also founded the company L.M. Ericsson.

The problem up till now was the one-to-one telephone line correspondence between subscribers. Therefore, the next important step was the development of the manual switching system to reduce the size of the mesh network. The first manual switching system was opened on 28 January 1878 in New Haven, Connecticut. The same year manual switches were installed in London and Paris. The International Bell Telephone Company installed the first manual switch in Kristiania (the former name of Oslo), Norway in June 1880.

The International Bell Telephone Company was established by Bell's father-in-law, Gardiner Hubbard in 1879. The company installed switches and access networks in several large cities in Europe. The European headquarters was in Antwerp where the company in cooperation with Western Electric built a large factory for telephone equipment. The company got a strong position, especially in Belgium, The Netherlands and Russia. The company charged their customers heavily and *prevented a natural evolution of the telephone penetration in these countries*.

In Kristiania, the Bell Company started by charging 100 NOK per year for a subscription, which corresponded to a two month salary for a telephone operator. The price was raised continuously and in the spring 1881 the subscription price was 200 NOK. The company received a lot of criticism. The prices were too high, the company did not cooperate sufficiently with the authorities regarding the telephone line tracks, and the building owners complained about the installers. The Government in Kristiania, as opposed to several other large European cities, had not given The International Bell Telephone Company sole rights for the telephone system. Therefore, Carl Söderberg and 12 businessmen from Kristiania founded *Christiania Telefonforening* on 24 May 1881. Söderberg had the L.M. Ericcson agency for



Manual telephone exchange (Bestorp, 1990)

telephone equipment in Norway, but telephone devices also from other telephone manufacturers were available. Carl Söderberg established in 1882 the independent Norwegian company, Elektrisk Bureau (EB); a competitor to L.M. Ericsson, which some years later had a yearly production capacity of 25,000 telephone sets and a considerable export.

Christiania Telefonforening's annual subscription price was set to 40 NOK, with 220 NOK for the installation. The Bell Company had to respond and reduced their subscription fee to 125 NOK, then to 100 NOK and later to 50 NOK.

The two competitors fought very hard to capture market share. The consequence was increased telephone penetration. The Bell Company did not succeed in getting sole rights in other Norwegian cities (Bestorp, 1990), (Rinde, 2005).

In 1885, the number of telephone subscriptions was 995 from Bell and 634 from Christiania Telefonforening. 230 subscribers had subscriptions in both networks. At that time the city government in Kristiania made it clear that the two companies had to merge because of the mess of telephone lines 'everywhere' and the possibilities for rationalisation and coordination.

The companies also got an ultimatum that they were not allowed to expand until a merger had taken place. The new company was the private stock company, *Christiania Telefonselskap*, which was established on 1 January 1886. A telephone monopoly was then established in Kristiania. However, the telephone subscription prices were on a reasonable level and the prices stayed constant for many years (Bestorp, 1990).

The International Bell Telephone Company, which originally was a threat to the telephone availability in Kristiania, had generated hard competition, low telephone prices and high demand. What happened in

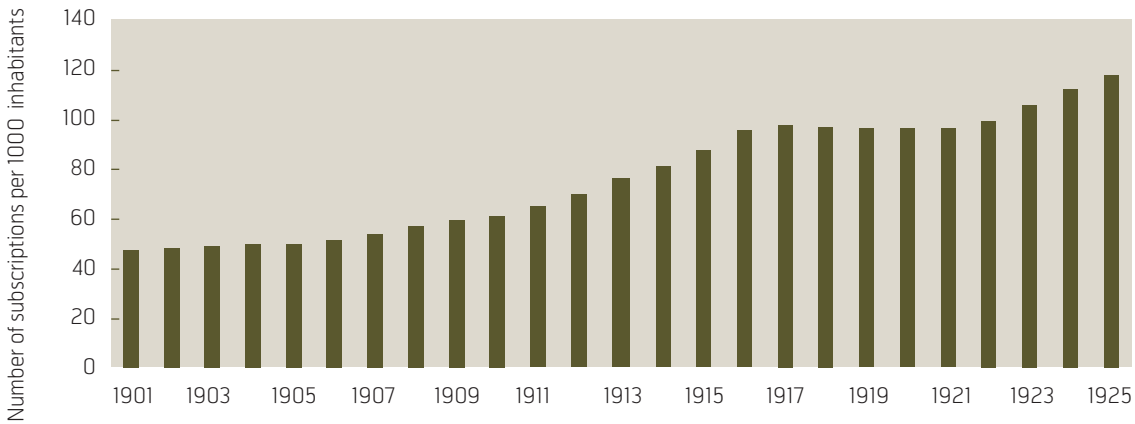
Sweden	15.6
Norway	15.0
Switzerland	12.4
Denmark	11.0
Germany	5.1
UK	5.1
Netherlands	3.3
Belgium	2.6
France	1.8
Austria	1.2
Spain	1.0
Hungary	0.9
Romania	0.3
Russia	0.3

Telephone penetration per 1000 inhabitants in European countries in 1900 (Rinde, 2005)

Kristiania also influenced the evolution in other places in Norway.

The table gives an overview of the telephone penetration in Europe in year 1900. The table shows that the Scandinavian countries together with Switzerland had the highest penetration. The penetration in Norway was 4.5 times higher than in The Netherlands and 5.8 times higher than in Belgium and about 3 times higher than the penetration in the well developed countries Germany and the UK.

The very special telephone growth in the Scandinavian countries during the first years could be called the *Scandinavian wonder* (Christensen, 2006). This observation is strengthened through the later mobile and broadband evolution in the Nordic countries. The Nordic countries have been pioneers in introducing



The evolution of the telephone penetration in Kristiania 1901 – 1925 (Bestorp, 1990)

ISSN 0085-7130

the Nordic Mobile System, NMT, already in the early 1980s, many years before other countries got the service. At the end of 2006 all five Nordic countries are among the eight OECD countries with the highest broadband penetration. As a conclusion all Nordic countries start very early to adopt new telecommunication technology.

In 1900 the number of subscribers in Kristiania was 9,864. There were 11,503 telephone sets, and each subscriber made in average 10.5 calls per working day. *A total of 27.8 million calls were carried through the main switch in Kristiania that year.* The telephone traffic was considerable (Bestorp, 1990).

Heavy investments in the national long distance network because of the increased traffic cleared the way for the start of a national telecommunication monopoly. The minister in charge of telecom, Jørgen Løvland, argued for the monopoly and in 1899 a law was passed giving Telegrafverket the exclusive rights to run telecom networks in Norway (Rinde, 2005), (Christensen, 2006).

Telegrafverket took over Christiania Telefonselskap on 1 January 1901. However, it turned out to be very expensive for the state to buy all the private telephone companies. As late as in the mid 1970s all private telephone companies in Norway were bought and embodied in Televerket (Telenor).

The first automatic telephone exchange in Norway was installed and operational in Skien in 1921 (Telektronikk, 61 (1-2), 1965, p 17)

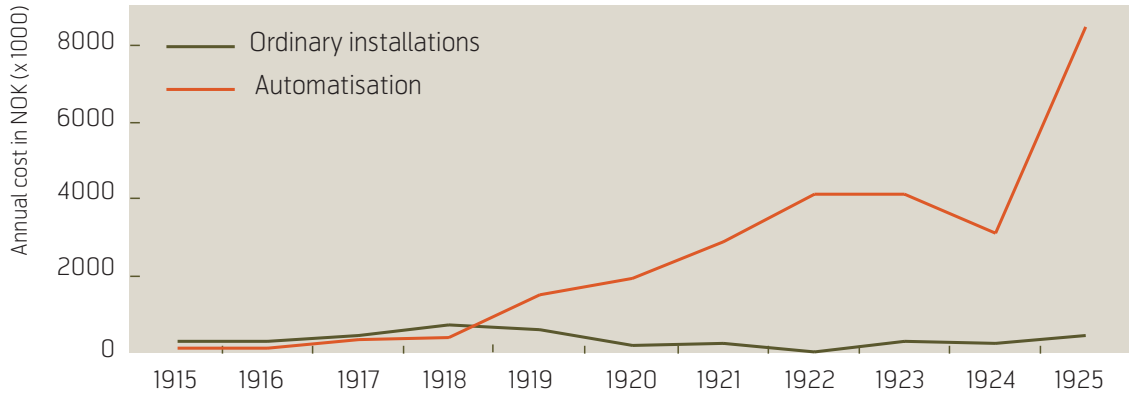
The subscription growth in Kristiania stagnated in the first years of the new century with only a few hundred new subscriptions each year. From 1907 the growth increased again and the total number of telephone subscriptions reached about 22,000 in 1913. During this period the government did not release sufficient investment means and problems with the traffic started to occur.

Start of the Automation

In 1914 only few subscribers could be connected to the telephone system in Kristiania. The situation was predicted several years earlier by Telegrafverket. The traffic increased and subscriber lines had to be moved from one switching group to another. Bestorp gives a detailed description of the situation in Kristiania caused by the increased traffic (Bestorp, 1990). A lot of subscriber complaints were received and the flood of complaints started in 1910. The subscribers were also irritated because the government used Telegrafverket as a money machine without giving back necessary investment means. The situation grew worse over the next years. From time to time parts of the network were totally blocked because of heavy traffic. In 1914 the Norwegian Parliament decided to increase the investments, but later lack of available equipment because of World War I limited the possibilities for expansion of the access network. The Ministry of Trade in cooperation with Telegrafverket appointed in 1918 a committee for withdrawing subscriptions from the subscribers. During a two year period 1,100 telephone sets were withdrawn in Kristiania and in the autumn of 1920 there were about 6,000 people on the waiting list. Copenhagen had about 5,000 people on the waiting list for telephone subscription.

This situation with considerable traffic problems and waiting lists for subscriptions was the backdrop for the pioneer work of Tore Olaus Engset and Agner Krarup Erlang. The development of traffic models for dimensioning of the switches and the access network was extremely important in a situation without sufficient investment means or available equipment.

The manual switches also had limitations. When the number of subscribers and traffic per subscriber increased, the capacity, including the number of telephone operators had to be increased. For a period the physical limits of the telephone company's premises prevented further expansion. And in July 1918 the Spanish flu hit Kristiania. Many telephone operators



Annual expenses for ordinary installations and automation in Kristiania 1915 – 1925 (Bestorp, 1990)

became infected causing significant traffic problems because of the lack of 'womanpower' at the switches.

In parallel, work was being done to start ordering new telephone switches. A three man committee with Telephone Director Iversen, Head of Department Engset, and Chief Engineer Abild was appointed in 1910. The committee's mandate was to give recommendations to the choice of a future switching system for Kristiania, either manual, semi-automatic or fully automatic. They spent 48 days travelling around Europe and 71 in the United States in 1911/1912. The recommendation was finalised in 1913, proposing fully automatic switching systems with primary and secondary exchanges. This envisaged a plan for 30,000 lines with a potential for 90,000 lines for Frogner exchange in Kristiania. The Norwegian Parliament sanctioned the plan in 1916 and Western Electric got the contract the same year. The first automatic exchange in Norway was to have been installed in 1917, but the ship *Kristianiafjord* transporting the exchange sank in June 1917. Because of World War I the project was delayed and the exchange was finally installed in 1921. At that time the private telephone companies in Skien and Bergen had already installed automatic exchanges, while a city like Stockholm still only had manual switches.

The figure above shows that investments for the establishment of automatic exchanges were very high in the first part of the 1920s. On the other hand, the investments also caused a significant reduction in the number of telephone operators. The number was reduced from 610 persons to 347 persons during the period 1924 – 1925.

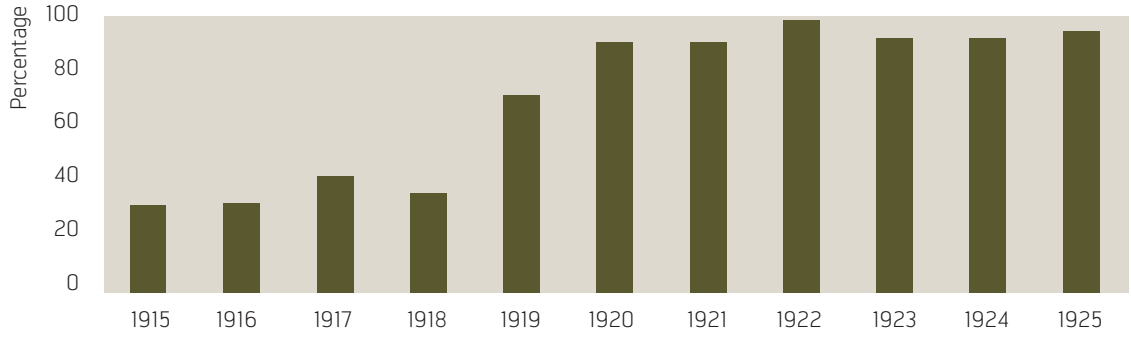
The figure below shows that the automatisation costs were completely dominating the other investment costs at the beginning of the 1920s.

Because of the very high investments, it was extremely important to have traffic dimensioning models which were fitted to the observed traffic. The very high investments in telephone exchanges underline the importance of having available traffic models and dimensioning tools.

Development of the Queuing Theory and the Teletraffic Theory

The Start of Work on Queuing Theory

Fr. Johannsen was appointed managing director of Copenhagen Telephone Company in 1903. He realized that the manual switches were not dimensioned in the right way. He published some pioneering work on this subject where he used the mathematical theory of probability (Johannsen, 1908 and 1910-11). From an economic point of view he stated:



Proportion of costs of the automation compared with total costs 1915 – 1925 (Bestorp, 1990)

ISSN 0085-7130

The overloading of the subscribers resulted in considerable extra expenses on account of the telephone operators having to make repeated attempts to establish a connection.

A comparison of the increased cost due to an extra line and the reduced costs of the telephone operators, since attempts were needed to establish a connection for different numbers of lines, gave a sound foundation for how many lines the company should require for the single subscriber. (Jensen, 1996)

To be able to develop more precise dimensioning Fr. Johannsen established a scientific laboratory where he engaged the mathematician Agner Krarup Erlang. Erlang started to work on the holding times in a telephone switch (Erlang, 1909) and he identified that the number of telephone conversations which satisfied a Poisson distribution as well as the telephone holding time was exponential distributed.

Tore Olaus Engset (1865 – 1943)

The next important step in the queuing theory was done by the Norwegian Tore Olaus Engset.

The life of Tore Olaus Engset is very well described in the book *Tore Olaus Engset – The man behind the formula* (Myskja, 2002). He was for a long period Head of the Administrative Department in Telegrafverket, which also contained a traffic unit. The position was directly below the General Director (DG). He also functioned as DG for a period in the 1920s. In 1930, the Government for the first time appointed a person inside Telegrafverket as General Director and the most natural choice was Tore Olaus Engset. He held that position until he retired in 1935. He was also honoured with the Commander of Second Order of Dannebrog and Knight of the Legion of Honour (Myskja, 2002).

Tore Olaus Engset (1865 – 1943)

It is very understandable that the primary work of this man during normal working hours was not queuing theory and traffic dimensioning models. As pointed out both in (Natvig, 2000) and in (Myskja, 2002), *"The work of Tore Olaus Engset is extremely impressive, partly because it was carried out in late night hours outside the traditional working hours and also because he did not have access to our modern mathematical theory of probability and numerical methods organised for computers"*.

As described in this paper traffic in the telephone systems grew significantly from 1910 onwards in Norway, and especially in Kristiania, causing overflow and blocking of calls in the networks. Hence, it was extremely important to make the right plans for extending the networks based on the available grants from the Government. Tore Olaus Engset's work was fundamental. He developed queuing models or teletraffic loss models for finite sources which could be used for dimensioning manual, semi-automatic and automatic telephone exchanges.

His approach to the dimensioning is extremely elegant! The documentation of the work is done through an unpublished 128 page report in 1915 – which was recovered by Villy Bæk Iversen in 1996 (Iversen, 1996). Engset also published his methodology in 1918 (Engset, 1918). His elegant approach, valid for many queuing systems, is mainly based on combinatorial modelling. He calculates the probabilities for having i lines in an exchange occupied, given statistical equilibrium, based on individual inter-arrival times and individual holding times for each of the N subscribers in the access area. The model is based on using all permutations in drawing i individual subscribers out of N. Then, Engset calculates the loss probability as a function of different sizes of K, the number of lines in the exchange. Of course, it is not very practical to do dimensioning based on different traffic characteristics of each subscriber, but it is very convenient to do so for different groups of subscribers where subscribers in the different groups have different 'traffic' behaviour.

Engset's methodology produces surprisingly general results, which up to this day are perfectly applicable in queuing theory (ITU, 2005), even if the methodology is not based on the traditional approaches used in queuing theory! Villy Bæk Iversen underlines that the model is insensitive to both inter-arrival distribution and the holding time distribution (Iversen, 1996).

A simplification of Engset's general model gives the well known Engset distribution, which is a truncated Binomial model. Here it is assumed that all subscribers have identical inter-arrival distribution with parameter λ, and all subscribers have identical service time distribution with the mean $1/\mu$. Let K be the number of lines and N the number of subscribers. Then Engset's distribution, where $p(i)$ is the probability for i occupied lines, is given by:

$$p(i) = \frac{\binom{N}{i}\left(\frac{\lambda}{\mu}\right)^i}{\sum_{k=1}^{K}\binom{N}{k}\left(\frac{\lambda}{\mu}\right)^k}$$

It should be noted that the famous Erlang B formula is a further simplification of Engset's distribution. Here, the number of sources is considered to be infinite, which of course is an approximation. Engset's simplified model is more precise because it assumes that the inter-arrival intensity, when i lines are occupied by the subscribers is $(M - i)\lambda$, while Erlang assumes that inter-arrival intensity is independent of the number of subscribers being serviced in the exchange.

Agner Krarup Erlang (1878 – 1929)

Agner Krarup Erlang (1878 – 1929)

Agner Krarup Erlang's life is well documented in (Brockmeyer, 1948). He finished his studies at the University of Copenhagen in 1901 acquiring the degree of candidatus magisterii (MA) with mathematics as principle subject. In 1908 the Copenhagen Telephone Company engaged Erlang. As pointed out he started to examine the holding times and published his first results in 1909 (Erlang, 1909).

Then in 1917, he published his most important work (Erlang, 1917). In section 1-7 he develops his famous Erlang B loss formula. As pointed out, the solution is based on considering equality of traffic streams through a cut between states.

Both Erlang and Engset have earlier been criticised when it comes to the validity of their models because they did not assume exponential holding time distributions. However, this criticism has been showed to be wrong, because the models are valid for different holding time distributions as long as there is a distinct mean (ITU, 2005).

Agner Krarup Erlang made a set of additional publications on teletraffic models in the 1920s and his famous loss formula, which was very applicable, got extremely popular for traffic engineers.

Erlang and Engset

This paper describes the incitements for developing teletraffic models. The telephone penetrations in the Scandinavian countries and Switzerland were significantly higher than in all other European countries at the start of the 20th century. However, from 1910 onwards the situation in Norway, and especially in Kristiania changed because of traffic congestions, a lot of complaints from the subscribers, limited investment grants, waiting lists for getting a telephone subscription etc. As mentioned earlier waiting lists were established – 5,000 potential subscribers were on the list in Copenhagen in 1920 and 6,000 in Kristiania. *This situation was the backdrop for the real start of development of the queuing theory. And the pioneers were Agner Krarup Erlang and Tore Olaus Engset!*

Erlang was employed by the Copenhagen Telephone Company in 1908 and Engset was already in 1894 Head of Traffic and Operations in Telegrafverket.

In 1910 Telegrafverket appointed a three man committee with Engset as one of the members to consider modernisation of the manual telephone systems in Kristiania by studying semi- and fully automatic telephone systems. The recommendation by the committee was finalised in 1913. During this period Engset had visited the Copenhagen Telephone Company and been aquainted with Erlang and P.V. Christensen who were active in traffic engineering. Hence, there are reasons to believe that Erlang and Engset exchanged views and knowledge on traffic modelling and dimensioning of the exchanges.

However, it is astonishing to realise that Erlang and Engset developed completely different approaches to

calculating loss probabilities and dimensioning. Both methods are excellent and history has shown that the methods are still 'future proof' (ITU, 2005).

We now know that Erlang's blocking formula is a simplification of Engset's model. A natural question is of course: Why is Erlang's work so well known and Engset's work is not?

Engset had for a long period developed his dimensioning method, which he documented (128 pages) in 1915. He sent the documentation to Copenhagen Telephone Company and probably also to Stockholm (Iversen, 1996). However, time elapsed and he did not get his work published until 1918 (Engset, 1918). The question is of course – why?

Engset was a very busy man. His main work was not teletraffic and queuing theory – even if these aspects were very important. The main thing for him could be only to develop the results, like Fermat did, and not use too much energy to publish his results.

Another reason could be the comments he got from Erlang via Fr. Johannsen, which made him aware of the fact that his formula is just an approximation to the model. Engset even quotes this as a footnote in his paper (Engset, 1918), (Jensen, 1992). The discussion was about offered traffic and carried traffic based on Engset's assumption regarding the observed traffic which he handles in the first part of his paper. This is a more general aspect, how to interpret the measurements and fit them to the modelling. The same arguments are also valid on Erlang's loss formula!

Exchange of Information

Nowadays it is easy to access scientific information. There is a number of sources like traditional text books, university courses, journals, libraries, search engines like Google, tailored conferences and of course, establishment of personal networks.

Search and exchange of information is carried out rapidly by using Internet and e-mail. Therefore, it is difficult to understand the situation 300 years ago or even 100 years ago when information was sent by letter. The most crucial point in old times must have been the accessibility to research and scientific information and especially information between different scientific areas.

Going back to (Struyck, 1716) and his solution of the Gambler's Ruin Problem, it is documented in this paper that he found the solution based on the same difference equations which 200 years later are used for solving the queuing system M/M/1. Engset used a completely different approach to solve the queuing systems, while Erlang used a similar approach. As mentioned in this paper, Erlang uses $i\mu$ instead of μ because of several servers instead of one. However, it must be pointed out that, while one thing is to solve the difference equations, another important thing is to deduce the equations itself.

Now, it is recognized that Erlang's B formula was immediately used by traffic engineers, while Engset's more general formula seems not to have been applied the first years. The main reason is probably that his solution simply was not known. Another reason could be that the formula is more complicated. The Engset loss formula admits the subscribers to have individual inter-arrival time distributions and holding time distributions. However, the simplification of his model assuming that all the subscribers have the same inter-arrival time and holding time distribution includes an additional parameter – the number of subscribers in the area compared to Erlang's B formula. Hence, the table becomes larger. At that time, there were no computers available for the calculations, so all relevant tables had to be produced 'by hand'.

Epilogue

This paper has briefly looked at the start of the mathematical theory of probability, the invention of the telephone and the start of the teletraffic/queuing theory. The following points have been drawn to attention:

Gerolamo Cardano was the real inventor of the mathematical theory of probability. However, he was not allowed by the Inquisition to publish his important work *Liber de Ludo Aleae – The Book on Games of Chance*. His work on probabilities was published in 1663, 87 years after his death, and inside a rather large ten cover volume of Cardano's publications.

Hence, Blaise Pascal and Pierre de Fermat are considered by many to be the real founders of the mathematical theory of probability. The exchange of letters between Pascal and Fermat in 1654 was known and had impact on the future evolution of the probability theory!

Andrew Wiles isolated himself for seven years to prove Fermat's last theorem. He published his proof, but there turned out to be a 'hole' in the proof. Happily, nobody managed to utilise the knowledge and Andrew Wiles completed the proof one year later!

Alexander Graham Bell has been considered to be the inventor of the telephone. This is not true. The Italian Antonio Meucci invented the telephone nearly 40 years before. He even tried to introduce the telephone

and have it patented in the American market, but did not succeed.

Then, Alexander Graham Bell obtained the patent and utilised all economic and commercial possibilities in an excellent way. Antonio Meucci did not get anything – except that the American Congress 113 years after his death declared that he was the real inventor of the telephone!

Agner Krarup Erlang and Tore Olaus Engset were the real founders of the teletraffic/queuing theory. Erlang developed his famous B loss formula in 1917. Engset developed a more general loss formula in 1915 based on a completely different approach, which also was sent to Erlang's company. Erlang's model is a radical simplification of Engset's model.

Erlang became very famous for his work, and his loss formula has been widely used by teletraffic engineers. Engset's work was rather unknown for a long period and has recently been appreciated.

History is complicated, but it shows that some are lucky and succeed, while others do not. Sometimes there are a lot of random elements which affect evolution. But, there may also be other factors. Regarding Pierre de Fermat and Tore Olaus Engset, they held important positions in society which at that time were given a higher priority than their more theoretical work.

References

(Bestorp, 1990) Bestorp, E. *Oslo telefonen 1880-1985* (The Oslo telephone 1880-1985). Televerket, Oslo district, 1990.

(Brockmeyer, 1948) Brockmeyer, E, Halstrøm, H L, Jensen, A. The life and work of A.K. Erlang. *Acta Polytechnical Scandinavica*, 2, Copenhagen, 1948.

(Christensen, 2006) Christensen, S A. *Switching relations. The rise and fall of the Norwegian Telecom Industry*. Oslo, Norwegian School of Management, September 2006. (Doctoral dissertation) (ISBN 82 7042 746 2)

(Cohen, 1957) Cohen, J W. The Generalised Engset Formulae. *Philips Telecommunications Review*, 18 (4), 158-170, 1957.

(Engset, 1915) Engset, T O. *On the Calculation of Switches in an Automatic Telephone System*. 130 pages report – Unpublished. Translated to English by A. Myskja in (Espvik, 2000).

(Engset, 1918) Engset, T O. De Wahrscheinlichkeitrechnung zur Bestimmung der Wählerzahl in automatischen Fernsprechämterm. *Elektrotechnische Zeitschrift*, Heft 31, 1918.

(Engset, 1921) Engset, T O. *Emploi de Calcul des Probabilitès pour la determination du nombre des selecteurs dans le Bureaux centraux*. RGE, Tome IX, 1921.

(Erlang, 1909) Erlang, A K. Sandsynlighetsregning og Telefonsamtaler (in Danish). *Nytt tidsskrift for Matematik B*, 20, 1909. Later published in French: Calcul des probabilités et conversations téléphoniques. *Revue général De l'Electricité*, 18, 1925.

(Erlang, 1917) Erlang, A K. Løsninger av nogle Problemer fra Sandsynlighetsberegningen av Betydning for automatisk Telefonsentraler (in Danish). *Elektroteknikeren*, 13, 1917.

(Erlang, 1918) Erlang, A K. Solution of some Problems in the Theory of Probabilities of Significance in Automatic Telephone Exchanges. *The Post Office Electrical Engineers' Journal*, 10, 1918.

(Espeli, 2005) Espeli, H. Det statsdominerte teleregime (The state dominated teleregime). *Norwegian Telecommunication history 1920-1970*, Volume II. Oslo, Gyldendal, 2005. (ISBN 82-05-334447)

(Espvik, 2000) Espvik, O, Myskja, A. *Telektronikk and the History of Engset*. Kjeller, Telenor R&D, 2000. (R&D N 78/2000)

(Hald, 1990) Hald, A. *A History of Probability and Statistics and their Applications before 1750*. Wiley, 1990.

(ITU, 2005) ITU. *Handbook Teletraffic Engineering*. Geneve, January 2005. (ITU-D, Study Group 2, Question 16/2)

(Iversen, 1996) Iversen, V B. Tore Engset – a Pioneer in Teletraffic Engineering. In: *Proc The Thirteenth Nordic Teletraffic Seminar*, NTNU, Trondheim, 20-22 August 1996.

(Jensen, 1954) Jensen, A. *A Distribution Model Applicable to Economics*. Copenhagen, Munksgaard, 1954.

(Jensen, 1992) Jensen, E. On the Engset Loss Formula. *Telektronikk*, 88 (1), 1992.

(Jensen, 1996) Jensen, A. Teletraffic theory: The Nordic School. In: *Proc The Thirteenth Nordic Teletraffic Seminar*, NTNU, Trondheim, 20-22 August 1996.

(Johannsen, 1908) Johannsen, F. Busy. Presented in: *The Development of Telephone Communications in Copenhagen 1891-1931*. Copenhagen, 1932. (Ingeniørvitenskapelige Skrifter, A No 32)

(Johannsen, 1910-11) Johannsen, F. Telephone Mangement in Large Cities. *The Post Office Electrical Engineers Journal*, (Oct 1910, Jan 1911). Reprint *Proc ITC 13*.

(Joys, 1967) Joys, L A. Til minne om T. Engset. *Verk og Virke*, 1967.

(King, 1963) King, A C, Read, C B. *Pathways to probability*. Holt, Rinehart and Winston, 1963.

(Kendell, 1953) Kendell, D G. Stochastic Processes occuring in the Theory of Queues and their Analysis by the Method of Imbedded Chain. *Ann. Math. Stat.*, 24, 1953.

(Kleinrock, 1975) Kleinrock, L. *Queueing Systems, Volume I, Theory*. New York, Wiley, 1975. (ISBN 0-471-49110-1)

(Kunnskapsforlaget, 2007) *Store Norske Leksikon*. Kunnskapsforlaget, Askehaug and Gyldendal. Internet version, Oslo, 2007.

(Mahoney, 1994) Mahoney, M S. *The Mathematical Career of Pierre de Fermat*. Princeton Univerity Press, rev 1994.

(Laplace, 1812) Laplace, P S de. *Théorie Analytique des Probabilités*. Paris. Reprinted in *Oeuvres*, vol 7, 1886.

(Morris, 1961) Morris, R, Wolman, E. A Note on Statistical Equilibrium. *Operations Research*, 9, 1961.

(Myskja, 1999) Myskja, A. T. Engset in New Light. In: *Proc ITC 16*, Edinburgh, 1999.

(Myskja, 2002) Myskja, A, Espvik, O. *Tore Olaus Engset – The man behind the formula*. Trondheim, Tapir, 2002.

(Natvig, 2000) Natvig, B. Køteori – en nøkkel til kortere ventetider. *Aftenposten*, 12 November 2000. Also at Forskningsdagene 2000.

(Ore, 1953) Ore, Ø. *Cardano, The gambling scholar*. Princeton University Press, 1953.

(Ore, 1960) Ore, Ø. Pascal and the invention of probability theory. *Amer. Math. Monthly*, 67, 409-419, 1960.

(Singh, 1997) Singh, S. *Fermat's Last Theorem*. Oslo, Aschehoug, 1997. (ISBN 82-03-20471-6)

(Stordahl, 1972) Stordahl, K. *Queing Systems in Statistical Equlibrium* (in Norwegian). Oslo, University of Oslo, 1972. (Post-graduate thesis)

(Stordahl, 2006) Stordahl, K. *Are the odds against you? The history of how gambling initiated the theory of probability and how the theory can be used to improve your odds*. Lecture for the dr.philos. dissertation, Norwegian University of Science and Technology, Trondheim, 19 December 2006.

(Struyck, 1716) Struyck, N. *Calculation of the Chances in Play, by means of Arithmetic and Algebra, together with Treatise on Lotteries and Interest* (in Dutch). Amsterdam, 1716. Reprinted in *Oeuvres*, 1-164, 1912.

(Todhunter, 1865) Todhunter, I. *A history of the mathematical theory of probability*. Macmillan and Co, 1865.

Kjell Stordahl received his MSc from the University of Oslo in 1972 and DrPhilos from the Norwegian University of Technology, Trondheim in 2006. He worked at Telenor's Research Department for 15 years – seven years as manager of the Teletraffic field. He was Chief Planning Manager in Telenor Networks from 1989 until 1996 and Manager of Market Analysis 1997–2002. Kjell Stordahl was appointed associated reporter and special reporter 1981–1988 in CCITT SGII for 'Forecasting International Traffic'. From 1985 to 1988 he participated in CCITT GAS 10 and developed a forecasting handbook. He has also worked for ITU's headquarters as a specialist on forecasting. From 1994 to 1997 he was on the Board of Telenor Consult AS. He was referee for Project Imagine 21 in the ESPRIT Programme 1999–2001. Kjell Stordahl was on the Technical Advisory Board of Virtual Photonics 2000–2002, and since 1992 he has participated in various projects funded by the European Commission: RACE/TITAN, ACTS/OPTIMUM, ACTS/TERA, IST/TONIC and CELTIC/ECOSYS where he has been responsible for working packages on market and demand. Kjell Stordahl has published over 160 papers in international journals and conferences.

kjell.stordahl@telenor.com