Part I

Introduction

The privacy debate is introduced, and it is explained how this thesis relates to it. The focus of this thesis is given: strong guarantees from cryptographic hash functions and formal analysis of security protocols. The structure and contents of this thesis are revealed. We present a case study of a potential and inspiring application domain: the police. We finish with some deliberations on the merits of central storage of sensitive information.

Chapter 1

Introduction

Securing sensitive information against unauthorized use, while guaranteeing access for its intended uses, is difficult. One of the difficulties lies in the fact that 'intended use' is a rather vague term, which is often explained as 'need to know', a similarly vague term.

Securing sensitive information in such a way that only specific, listed people have access to it, is not that difficult. For example, many police records are protected in such a way that police officers have access to it, but others not. The problem with this type of protection is, that police officers do not need access to *all* police records, but only to those *relevant* to them. Though police officers are generally trustworthy people, they do have access to information they have nothing to do with.

Whether there is a need to know, seldom coincides with a list of people that is easy to construct. If one wants to improve the guarantees against unauthorized use of sensitive information, it seems logical to grant access to information based on whether there is a need to know. The problem is that it is difficult to precisely state and operationalize what constitutes a need to know. Moreover, the question whether one should be granted access, often depends on the information itself, to which there is no access yet: a Catch-22.

It is the aim of this thesis to help solve this problem, by offering technical solutions: we propose new *methods* to organize information in databases, and we propose cryptographic *protocols*.

- Using the *methods*, it is possible to compartmentalize the information into pieces that are so small that the need to know is easier to operationalize.

- Using the *protocols*, it is possible to make access decisions based on secret information without disclosing the information.

1.1 The Privacy Debate

Informally, 'sensitive information' is information about individuals or organizations, for which there is some commonly accepted norm that not all that information should be available to everybody. Typically, this is in the interest of the individuals or organizations that the information is about. The aim of the confidentiality of sensitive information is to prevent misuse of the information.

The trouble is, that for many kinds of sensitive information, there are genuine reasons why in certain circumstances, the confidentiality has to be sacrificed. When one wants to facilitate such 'genuine sacrifices', there needs to be a 'backdoor' in the protection of the information. The appropriate way to implement this proverbial backdoor is the subject of a large public debate, the *privacy debate*.

The following examples help to get an impression of the kinds of sensitive information to which the privacy debate applies:

Criminal records Information about the criminal behavior of people is kept secret in order to guarantee the state monopoly of justice. By keeping criminal records confidential, it is prevented that people take the law into their own hands or interfere with their own prosecution, which would result in disproportional and unequal punishment.

Medical files Information about the particulars of someone's health is kept secret as an extension of the integrity of the body. The medical history of someone is considered as very intimate information. Keeping medical files confidential promotes solidarity. If medical files were public, persons who fall into risk categories of whatever kind, would have great difficulty in applying for jobs and obtaining insurance.

Mail The confidentiality of mail is there to guarantee the freedom of thought and speech. If one has no control over who will receive a mail message when it is sent, one cannot verify whether the recipient is a friend or possibly a foe. Lack of confidentiality of mail will lead to some form of self-censorship.

Financial files Details of the financial status of an individual are kept secret to protect the negotiation position of individuals, and to protect wealthy individuals from targeted crime. Confidentiality of financial files should prevent exploitation.

The underlying norms that make this information 'sensitive' are not controversial, they are not the subject of a serious debate. Nevertheless, it is also commonly understood that the confidentiality of this information cannot be absolute. For certain jobs in which people carry a high responsibility, the criminal records of applicants have to be verified. If someone carries a highly contagious and dangerous disease, health organizations will use this information in order to prevent an epidemic. When it is certain that a person is planning a terrorist attack, his mail will be read and his phone tapped, in order to prevent

the attack. When someone goes bankrupt, the curator will have full access to that person's financial files.

There is a fierce public debate about at what point the interests of the society as a whole should take precedence over the interests of the individual. For example: how strong should evidence be before the privacy of an individual is infringed? If the evidence is required to be 100% decisive and based on *sound science*,[1] the interest of the society will in practice almost never take precedence. If the requirements on the evidence are set too low (say, only 1%), one faces the risk of sacrificing the fundamental values of justice [Sus06]. Here is a Catch-22: if the evidence is in fact 100% decisive, the information is probably not even needed anymore, but to know that it is indeed 100% decisive, one needs access to the information.

This *privacy debate* is widely believed to have zero-sum characteristics[2]: the wishes of those who defend privacy are (supposedly) fundamentally incompatible with the wishes of those who give priority to fighting crime and terrorism. The thought that describes this, can be summarized as 'Either you infringe everyone's privacy, or you do not catch any terrorists.'

In this thesis, the author does not wish to engage in this debate. Its connotation is primarily political and judicial, and though the author is well informed on these subjects and has a stake in this debate, he does not wish to claim scientific expertise on them. The opinion of the author is a *political* one, and not one *based on science*. Personal opinions do not belong in a scientific work.

Whatever comes out of the privacy debate, whether it be by means of consensus or decree, it will be a policy that tries to reconcile confidentiality of sensitive information with the guarantee of its use in urgent circumstances. When such a policy is to be implemented, scientific questions arise, such as:

- How can we organize information in such a way that a policy is enforced?
- What kind of guarantees on confidentiality and availability are precisely required?

It is not uncommon that a chosen policy cannot be implemented as such, but has to be adjusted before it can be implemented: not everything that is desirable is also possible (or affordable). Adjustment of a policy prior to implementation may mean that delicate agreements reflected in the policy are sacrificed for the sake of being able to implement the policy at all. A good policy is one which can be implemented without sacrificing essential parts of the policy. As such, technology has a *constraining* effect on policy-making, while technology should ideally be *facilitating* for policy.

[1] 'Sound science' is a political euphemism for requiring a very high burden of proof [Moo05].

[2] To see the zero-sum characteristics, it helps to perceive the privacy debate as a game between the privacy advocate and the terrorist fighter. The goal of the game is to settle on an information exchange policy. From the established policy, player payoffs are calculated. If infringing privacy is required for catching terrorists, the interests of the players are (at least to some extent) diametrically opposed [Bin92, page 237] [OR94].

The aim of this thesis is to increase the *facilitation* of technology for policies regarding the exchange of sensitive information. In that perspective, the key contributions of this thesis are the following:

Information Designators A new method for structuring information, which demonstrates that exchanging information on the one hand, and privacy and confidentiality on the other hand, can go hand in hand (Chapter 7).

Knowledge Authentication A set of efficient protocols which allow the comparison of information without disclosing the information itself. This has applications in, for example, passenger information exchange between the European Union and the United States of America (Chapters 8–10).

Thus, although this thesis does not defend any position in the privacy debate, it provides input to this debate. The techniques presented in this thesis offer new solutions to settle the privacy debate. The relevance of these techniques is high because they demonstrate that in certain cases it is possible to simultaneously accommodate many wishes of either side of the privacy debate: to respect the privacy of innocent citizens to a high degree, while the information required for protecting the society as a whole (against terrorism, contagious diseases, etc.) is provided.

It is not to be expected that any reasonable civil liberties activist will object if the privacy of proven criminals or terrorists is infringed on purely for the act of bringing them to justice[3]. We present technologies that do just that, without infringing on the privacy of innocent citizens. Thus, we demonstrate that settling the debate is not a zero-sum game at all if the right technology is deployed.

1.2 Guarantees of Availability and Confidentiality

The meaning of a 'guarantee' in a legal context is quite different from the meaning of a guarantee in a mathematical or cryptographic context. This applies also to guarantees of availability and confidentiality.

A legal guarantee of *availability* of information, such as the *Freedom of Information Act* (FOIA)[4] gives someone a right to certain information, but not a means to quickly exercise this right. A FOIA request may take months to complete, and the request may fail for a number of reasons.

A legal guarantee of *confidentiality* of information, such as a *privacy law* or a *non-disclosure agreement* (NDA) obliges someone to keep certain information secret, but it does not physically prevent disclosure of the information. An

[3] Because it implies the information collection procedure is selective [Jac05].

[4] This is the name of two similar laws in the United States of America and the United Kingdom. The Dutch equivalent of this law is the *Wet Openbaarheid Bestuur* (WOB). The FOIA is best summarized as "The Freedom of Information Act gives everyone the right to access information held by the public sector." (http://www.direct.gov.uk/RightsAndResponsibilities/RightsAndResponsibilitiesArticles/fs/en?CONTENT_ID=4003239&chk=xi42h7)

NDA gives the owner of the information the right to hold the discloser of the information liable in court. This leaves the owner of the information at risk: he may be unable to detect the disclosure of the information, while he does suffer from damages as a result of the disclosure.[5] Even if he is able to detect the disclosure of the information, he may be unable to prove this convincingly in a courtroom setting. Not to mention that going to court is generally a tedious, time-consuming and expensive process.

While legal guarantees of availability and confidentiality effectively serve a large number of purposes, such as the protection of press freedom and intellectual property, they are often considered insufficient for parties at either side of the privacy debate. A police officer cannot do his duty if he depends on time-consuming procedures to obtain the information he needs to investigate a crime. Similarly, a fighter for civil liberties will not be satisfied with a mere *promise* that his phone will not be tapped, as is it highly unlikely that he will ever be informed of such a tap.

Cryptography is not like law at all. One of the aims of cryptography is to guarantee information availability and confidentiality to absurdly high degrees.[6] Instead of writing lengthy contracts in legalese that describe what should happen when something goes wrong, cryptography aims to *prevent* things from going wrong. For example, to break the best modern encryption, which is a means to guarantee confidentiality, one needs thousands of years of computation. Colloquially, the likelihood of guessing an encryption key correctly is smaller than the likelihood of getting struck by lightning several days in a row, and surviving.

Cryptography and modern IT infrastructure form a tandem which has the potential of offering both confidentiality and availability guarantees. Admittedly, IT infrastructure does not always live up to its expectations, as it sometimes fails to deliver availability – also in cases where confidentiality is of no concern. If cryptography is improperly used, it does not offer any confidentiality guarantees.

This thesis aims to help develop the potential of the cryptography/IT tandem for guaranteeing both confidentiality and availability of sensitive information.[7] In this perspective, we focus on the following:

Cryptographic Hash Functions A type of algorithm for creating *fingerprints* of information, in such a way that the information itself is kept confidential. Our contribution is a new application area of cryptographic hash functions, and the concept of a *non-incremental hash function* (Chapter 3 and throughout the thesis).

[5] This risk is not hypothetical: for example, it is not unusual that a creditcard or mortgage company refuses someone as a client based on information gathered from an information broker (a 'mass-scale private investigator') that gathered information from obscure sources.

[6] Other typical aims of cryptography are non-repudiation and guarantees of integrity.

[7] Digital Rights Management (DRM) technology is similar but different. DRM technology guarantees confidentiality and availability of copyrighted material, which is typically owned by the entertainment industry. The goal of our research is to promote both confidentiality and availability of sensitive information about individual citizens.

Authentication Logics A method for analyzing security protocols. Our contribution is the extension of a particular authentication logic (GNY logic) to handle cryptographic hash functions, and a proof that another authentication logic (BAN logic) improperly handles cryptographic hash functions. (Chapters 4 and 5)

This thesis focuses on strong guarantees for availability and confidentiality. We use formal methods to reason about the knowledge of persons involved in security protocols, and analyze what these persons can and cannot derive from the information they have got.

1.3 Thesis Contents

In the previous pages, we have motivated and introduced the research that is presented in this thesis. The structure of the thesis at hand is as follows:

In Part I we motivate the relevance of the research conducted in this thesis. What is sensitive information, and why do we need technology to handle sensitive information? (Chapter 1). For reference and the uninitiated, we summarize some important concepts of security and cryptography (Chapter 2).

Part II introduces the main security building blocks used to construct and analyze the protocols presented in this thesis. The *cryptographic hash function* is the most important cryptographic primitive applied in this thesis. We introduce the concept of a *non-incremental* cryptographic hash function (Chapter 3). To analyze the security properties of the presented protocols, we use *authentication logics*, of which the basics are explained (Chapter 4). BAN logic, the 'mother of all authentication logics', contains a flaw with how it handles cryptographic hash functions. We prove that as a result of this, BAN logic is not 'sound' (Chapter 5). GNY logic, another authentication logic (summarized in Appendix B), is extended in such a way that our protocols can be analyzed using GNY logic (Chapter 6).

Part III is the conceptual heart of the thesis, and presents two approaches to handling sensitive information. Instead of propagating information throughout information systems, which is essentially a massive disclosure of information, information can also be linked by using *information designators*. This idea may sound trivial, but it is in fact a radical departure from how information is integrated nowadays. If the idea is fully applied, there are benefits for both information availability and confidentiality (Chapter 7). A typical problem with confidential information is how to establish that two people both know something without actually disclosing it, which we dub *knowledge authentication*. The chicken-and-egg problem 'do you know the secrets that I know?' turns out to be difficult to formalize. We distinguish the case *1-to-many*, where one single secret (that I know) is compared to many secrets (that you know) and the case *many-to-many*, where many secrets (that I know) are compared to many secrets (that you know). We survey what solutions currently exist for this problem, and what their merits are (Chapter 8).

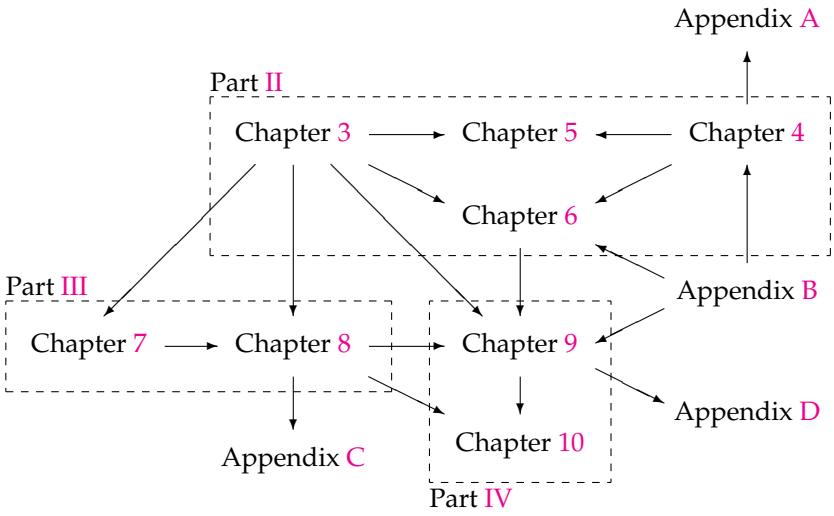

FIGURE 1.1: Dependencies between the chapters that make up the main body of the thesis at hand. The arrow $y \longrightarrow x$ means that to *fully* understand chapter x, chapter y should be read.

In Part IV, new security protocols are presented which implement knowledge authentication. The T-1 protocol does so in the 1-to-many case. In the T-1 protocol a message is sent that can only be recognized if one knows a particular secret. When one claims to know the secret, a (NP-hard) puzzle is created that can only be solved if one knows the secret. The T-1 protocol is very efficient in terms of communication and computational requirements. It uses non-incremental cryptographic hash functions, and is proven correct using our extension of GNY logic (Chapter 9). The T-2 protocol implements knowledge authentication in the many-to-many case. The T-2 protocol is a parallel composition of the T-1 protocol, and as such inherits its security properties from the T-1 protocol. In the T-2 protocol the players work together to efficiently determine the intersection of their secrets. The communication complexity of the T-2 protocol depends on the cooperativeness of the players, and we estimate this complexity experimentally (Chapter 10).

In Part V, we draw some conclusions (Chapter 11).

The appendices in Part VI contain some background information to various chapters. Appendices A and C contain some in-depth explanations to the literature referenced in Chapters 4 and 8, respectively. Appendix B gives a summary of GNY logic which is used in Chapters 4, 6 and 9. Appendix D describes the prototype software of the T-1 protocol presented in Chapter 9. Appendix E gives a summary of the formal notations used in this thesis.

When we focus on the main body of the thesis at hand (leaving out the introduction and conclusion part) we can visualize the dependencies between

the chapters as in Figure 1.1. When chapter x depends on chapter y (notation $y \rightarrowtail x$), this means that to *fully* understand chapter x, chapter y should be read. For Chapters 7 (information designators) and 8 (knowledge authentication), special care has been taken to make sure they are still rather intelligible without reading their background knowledge chapters first (Chapters 2, 3 and 7). In Chapter 9, it is only the correctness proof in GNY logic of the T-1 protocol, given in Section 9.4, that depends on Chapter 6.

Chapters 3 (cryptographic hash functions) and 4 (authentication logics) introduce the background on which the results in this thesis are based. As such, these chapters mainly summarize results of other researchers, though some new concepts are introduced in these chapters: *non-incremental cryptographic hash functions* (Chapter 3) and *heavy GNY annotations* (Chapter 4). From Chapter 5 onwards, all material presented is essentially new, though Chapters 7 and 8 contain various discussions of related research.[8]

1.4 Relation to the Author's Other Publications

Chapter 5 has been presented at FAMAS'06 [Tee06a]. An earlier version of Chapter 7 has appeared in the International Journal of Computer Systems Science & Engineering [Tee05b].

The T-1 protocol which is presented in Chapter 9 was first published at the Knowledge and Games Workshop 2004 [Tee04], and its proof in GNY logic (Section 9.4) first appeared in Synthese [Tee06b]. Chapter 8 (in particular Section 8.2) contains traces of both of these publications. Our opinion letter in Het Financieele Dagblad [Tee05c] can be considered a summary for laymen of the T-1 protocol. All our publications mentioned above are subsumed by this thesis.

The author's publications on workflow analysis and security [TvdRO03, TvdRO02, Tee99] and on expert systems for online voting advice ('party profile websites') [HT07, Tee05a, TH05] are not reflected in this thesis.[9]

1.5 A Case Study: the Dutch Police

A typical example of sensitive information that has to be kept confidential, is the information maintained by crime investigators of the police. Dissemination of this information to criminals would render the research of those police officers almost worthless. At the same time, it has to be prevented that two or more research teams investigate the same people without knowing this. When police teams are unaware of such a situation, their actions can easily harm one another's research. For example: one team is shadowing a suspect in the hope

[8] Moreover, Appendices A, B and C do not present new work, but contain discussions of related research. Appendix D can be considered new work, as it presents a prototype of our new T-1 protocol.

[9] The publications mentioned here do not constitute an exhaustive list.

that his actions will reveal new pieces of evidence; now another team runs in and arrests the suspect for another crime. If police teams cooperate, such situations can be prevented, and moreover collaboration of teams may help them to mutually exchange incriminating evidence.

In this section, we will describe how the Dutch police currently handles this situation and tries to reconcile information exchange and confidentiality.[10] It shows the current practice of how an organization of professionals with a high responsibility deals withs sensitive information.

It should be noted that there are many more information exchange systems operational within the police, for several purposes (criminal records, missing persons tracing, fingerprints, license plate registrations, etc.). The system described is just one of the systems deployed by the Dutch police.

For every investigation project performed by the police, an electronic dossier[11] is created. Access to this dossier is only granted to the officers dedicated to the project, and a number of superiors. Such dossiers are created and maintained locally, at a local police department[12], and can have any subject, varying from 'the great train robbery' to 'pickpockets in Amsterdam'.

An electronic dossier consists of records[13] which are added to the dossier over time. Every record has a time stamp. A record consists of various fields[14], in which the actual information is stored. There are fields for plain text, but also fields to store specific types of information. The most important types are:

BTK for persons[15],

ORG for organizations,

LOK for locations, such as physical addresses,

COM for means of communication, such as phone numbers, and

VTG for vehicles[16].

For all of these five field types, there are instructions on how the information should be encoded. For VTG, the encoding is simply the license plate number, for other types, the encoding is more complicated. For BTK for example, there is a method to derive a 20-character string from a name and birth date, of which it is supposed that it uniquely identifies any person.

[10] Interview with Tom van der Velde, privacy officer of the Police in Groningen, conducted on November 28, 2002, together with Pieter Dijkstra. The information has been verified by Paul Elzinga, IT specialist of NPOL, the organization which manages the information exchange infrastructure of the Dutch police. The interview occured in the context of the ANITA project, which was supported by the Netherlands Organisation for Scientific Research (NWO) under project number 634.000.017.

[11] In Dutch police jargon, an electronic dossier is called a *registratie*.

[12] In Dutch police jargon, this is at the *korps* or *regio* level.

[13] In Dutch police jargon, a record is called a *mutatie*.

[14] In Dutch police jargon, a field is called an *object*.

[15] BTK stems from *betrokennen*.

[16] VTG stems from *voertuigen*.

Hopefully, the police officers will nicely follow the instructions and fill in all relevant places in the prescribed way. As with every database which is manually filled, the electronic dossiers also suffer from inconsistent annotation, and use of text fields where the specific fields could have been used.

When the electronic dossier has as subject a crime for which it is possible to obtain an arrest warrant[17] and it is to be expected that the investigation will take more than one week, there should be a record in the electronic dossier which is tagged MRO[18]. MRO records can only contain fields of the types BTK, ORG, LOK, COM and VTG. Electronic dossiers can be grouped into two classes: those that contain an MRO record[19], and those that do not contain an MRO record[20]. The former class is typical for investigations of serious crimes, the latter class is typical for exploratory investigations. There are some rules about what information should be filed in an MRO record, and what information should not be filed in an MRO record. Roughly, *primary* information about the investigated crime, such as the suspect and the crime scene location, belongs in an MRO record, and *secondary* information, such as the phone numbers found in the agenda of the victim do not.

There is one very special electronic dossier, which is called ZWACRI[21]. It is essentially a black list of known 'professional' criminals. In the ZWACRI dossier, only the fields BTK and ORG are allowed.[22]

To improve the individual investigations, the electronic dossiers are automatically compared. The system that performs the comparison is called the VROS[23], and is housed in a heavily protected bunker which would not look misplaced in a James Bond movie. Every local police office has a dedicated, encrypted data connection to the VROS, and every Sunday morning, all electronic dossiers which contain an MRO record and the ZWACRI dossier are transmitted to the VROS. Only fields of the types BTK, ORG, LOK, COM and VTG are transmitted; text fields are omitted.

In the VROS, the information is divided into the following groups:

MRO which contains the MRO records,

Black Box which contains records that are not tagged MRO,

ZwaCri which contains the ZWACRI dossier, the black list of known 'professional' criminals.

---

[17] i.e., for which arrest and detention is allowed, in Dutch *voorlopige hechtenis*, artikel 67 Wetboek van Strafvordering.

[18] MRO stands for *Melding Recherche Onderzoek*, which translates roughly to 'notice of police investigation'.

[19] In Dutch police jargon, these are called *Lopende Recherche Onderzoeks Registraties* (LROR).

[20] In Dutch police jargon, these are called *Aandachtsvestigingsregistraties*.

[21] ZWACRI stems from *zware criminaliteit*.

[22] Within the ZWACRI dossier, there are two groups of listed people: the 'S' group (for *subject*) contains those who are permanently on the black list, and the 'V' group (for *voorlopig*), who are only temporarily on the black list. The 'V' group is also called *grijze velders*. The ZWACRI dossier is also called the *S/V-index* or *Subjecten-index*.

[23] In Dutch, this stands for *Verwijsindex Recherche Onderzoeken en Subjecten*.

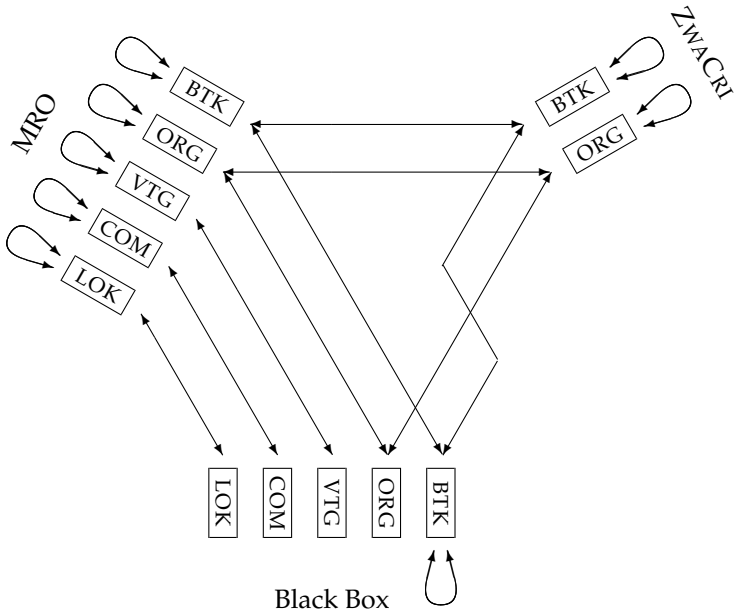

FIGURE 1.2: Matching of police information within the VROS. Every arrow represents a comparison (matching) procedure.

Within every group, all information of every field type (such as COM) is searched for duplicates (except in the Black Box, where this search is only performed for BTK fields). When in two groups the same type of information exists (such as COM), these groups are also searched for duplicates.[24] When a duplicate (a match) is found, the owners of the corresponding electronic dossiers are informed of this, with one another's contact information. Thus, every Monday morning, many police officers call one another with questions like 'I heard you are also investigating Willem Holleeder, could we find out if we can share some information?'. From there on, the process of information exchange depends on the sagacity and discretion of the police officers. The comparison that is performed by the VROS is best depicted by Figure 1.2.

After the comparison process is carried out, the information resident in the Black Box group is deleted. The information in the MRO and ZWACRI group are retained until the next Sunday morning, and is available for queries by police officers. During the week, this information at the VROS is not updated. The next Sunday morning, the process starts all over again. When a duplicate (a match) is found where both copies stem from a record which is more than

[24] Note that this matching process handles the information in non-MRO records differently from information in MRO-records: matches between two non-MRO records are only found when it involves a person. Thus, *secondary* information is only mutually compared if it involves a person. *Primary* information of any type is compared with all other primary and secondary information.

one week old (which can be seen from the time stamp), the match is no longer reported. This prevents matches from being reported more than once.

To summarize, the VROS can be consulted in two ways. The first way is implicit: by just adding records to an electronic dossier, matches are reported once every week. The second way is explicit: by typing in a search query during the week. One can only search for exact matches, queries like 'give me all people suspected of or convicted for blackmail' are not supported.

1.6 Considering Central Storage

The information exchange in the police organization, which we described in the previous section, is a sophisticated system. Sophistication should not be confused with quality of protection. There are a number of weak spots in the system.

First of all, the information that resides at the VROS is unencrypted. A small group of insiders with access to the 'James Bond' bunker in which the VROS resides, can effectively read, use and distribute the information of the VROS at their own will. This is 'the risk of the malicious system administrator'. Because the system is large and the stakes are high, the potential damage is huge. The concentration of information in the VROS makes it an interesting target for crime, whether it be in the form of corruption (bribes) or attack (computer break-in).

Secondly, the VROS is not protected against frivolous, unnecessary queries. Any police officer that performs investigations has access to the VROS. Any police officer can check whether relatives and celebrities are subject of investigation, for example. Because the number of police officers is large, and the corpus of information they have access to is large, dissemination of the information can almost be taken for granted.[25]

The central problem with the quality of the protection of the police information in the VROS boils down to central storage of information, and too generous access to the central information.

The 'VROS solution' does not scale up internationally. Police organizations are often willing to assist their sister-organizations in another country, but they do not unconditionally trust one another[26]. Moreover, the law would probably prohibit such information exchange in various ways. Thus, central storage of police information is not feasible on an international scale.

These problems are not unique for the Dutch police, but apply to all sensi-

[25] Another police information system where this is very clear is the Dutch license plate registration, to which officially only the police and some selected government organizations have full access. In practice, it is rather easy to collect this information for a given license plate number.

[26] Within a single country, the police departments are essentially obliged to trust one another by the central government. In an international context, such a central authority is lacking (by definition).

tive information held by the government. Let us take for example the discussion about the Dutch BSN[27].

The Dutch government maintains a lot of information about individual citizens, such as their employment status, marital status, enrollments in educational institutions, possession of real estate, etc. When one applies for welfare (financial aid) one has to provide this information to the government, while the information is already within the possession of the government. This situation creates an unnecessary administrative burden for the citizen, and also a possibility for fraud [HS06]. Following this line of reasoning, one might conclude that a central storage of the information about individual citizens, not much unlike the VROS, linked by means of the Dutch BSN can prevent fraud and increase service of the government.

The criticisms against the BSN are very similar to the central problems of the VROS, identified above. Mommers[28] writes [Mom06][29]:

> "The problem is, that for proper use of the BSN, we have to assume an outright incorruptible government, which will never use information for other purposes than those which are strictly necessary, and which will never undesirably give the information to others. Moreover, we must assume that it is desirable that the correct information is available always and in every circumstance."

The argument of the incorruptible government ('risk of the malicious system administrator') also holds into the future, as long as the information is retained. One may trust the current government, but maybe not the future government. In the second world war, the Nazis gratefully used the Dutch municipal inhabitants register, which recorded for every citizen the religious affiliation, including the Jews.[30] Alberdingk Thijm[31] can be considered more than just critical, and mentions a somewhat similar scenario [SOL06, page 71]:

> "Privacy violations will be omnipresent the coming years. The passed few years [sic], it was not considered decent to stand on the barricades for the right to privacy. This is largely due to 9/11 and the cry for more anti-terror legislation and less privacy-protection. The department of justice recently set up a system that enables civilians to participate in crime-fighting by recording criminal events they witnessed with mobile phones. It is striking that when such a system is proposed, the Data Protection Board[32] does not object.

[27] BSN stands for *Burger Service Nummer*, which translates into 'citizen service ID'. It is comparable to the American *social security number*.

[28] Associate professor at eLaw@Leiden, who performs research on accessibility of legal information and the interaction between law and technology.

[29] This quote is translated from Dutch by the author, and authorized by Mommers.

[30] The underground resistance eventually identified this problem, and started destroying municipals registers. The Amsterdam municipal register was set on fire on March 27, 1943.

[31] Alberdingk Thijm is also known as the attorney of Kazaa and Skype.

[32] College Bescherming Persoonsgegevens — WT

> It is obvious that there are many problems attached to such a system, such as the authenticity of the material sent by the civilians, access-control and the duration of the storage of it. Only when the supermarket bonus card[33] administration is hacked by Al Qaida [sic], or something similar, civilians will start worrying about their privacy again."

All in all, there is a large number of reasons why central storage of sensitive information is not advisable or desirable [Jac05]. To guarantee the availability of sensitive information where it is needed, *some form* of central storage is required. But is it required to store the sensitive information *itself* in a central location? Can we find a solution in which only *references* to the sensitive information are stored centrally?

The VROS was designed with these considerations in mind: the electronic dossiers are not *completely* sent to the VROS, and the VROS offers only *pointers* to other electronic dossiers and their contact persons.

A similar thing is now happening with medical files in the Netherlands: there is an index of medical files, the LSP[34] [Spa05], which links medical files to persons based on their BSN. As a result, though the LSP by itself does not contain any medical file, it is rather trivial for almost anybody in the health business to estamate the size of a persons medical record, which is of course a proxy for the general health of the person in question[35]. This demonstrates that a centralized index which does not contain any 'sensitive file' by itself, but which does contain links to such files, may still disclose information that should be better protected.

In this thesis, we present technical solutions (the T-1 and T-2 protocols, Chapters 8–10) that allow the VROS to do exactly what is does now, but without storing the sensitive information in a legible form, as it does now. Thus, the risk of central storage ceases to exist.[36] The same techniques can be used for comparing bodies of information when there is no central authority, such as the exchange of passenger information exchange between the European Union and the United States of America.

Similarly, the *information designator* (Chapter 7) is a technique that guarantees availability without creating central storage of sensitive information.

[33] Loyalty program of Albert Heijn, a large supermarket brand in the Netherlands — WT

[34] In Dutch, this stands for *Landelijk SchakelPunt*.

[35] Note that though the BSN of a person is officially secret, it is really easy to obtain this number, as it is printed on many documents, such as the drivers' licence and the passport.

[36] Moreover, this may save the Dutch police a few pennies because the need for the 'James Bond'-bunker ceases to exist.