# Advanced Applications for e-ID Cards in Flanders

## ADAPID Deliverable D11

## Basic Research II

C. Diaz (Ed.), B. De Decker, M. Layouni, G. Nigusse, B. Preneel, A. Rial, C. Troncoso, B. Van Alsenoy, H. Vangheluwe, and K. Verslype

Version 1.0

February 2009

# Executive Summary

This report presents the basic research results of the ADAPID project from in 2007 and 2008. It summarizes twenty research papers, most of which have been published in scientific journals of conferences, while the rest are still subject to submission or review processes. Our research results can be classified in five main areas, corresponding to each of the chapters in this report.

First, we present our work on digital credential systems that provide special security and privacy properties. These digital credentials are a key technology for building privacy friendly user-centric identity management systems. Our results improve the flexibility, efficiency and functionalities offered by these digital credentials. Moreover, we show how these credentials can be useful in e-health systems.

Second, we describe our contributions to the formalization of privacy properties, and propose improvements to both information theoretic as well as combinatorial anonymity metrics. We also present an evaluation of mix based communication systems by proposing a new attack methodology that is far more effective than previous work, and further demonstrates the difficulty of achieving privacy when the communication layer is taken into account.

Third, we introduce a diverse set of techniques to fulfill security goals for data storage and retrieval. The research problems tackled include secure long-term archiving, negative databases, steganographic file systems, private data search, private data retrieval, and priced oblivious transfer.

Fourth, we present our work on secure and privacy enhanced applications, using the e-ID as basic authentication token. The researched applications include electronic petitions, pay-as-you-drive insurance systems, electronic ticketing, and credential vaults. We also include a study on the integration of biometric authentication in e-ID cards.

Finally, we analyze from a legal perspective the privacy risks presented by the Belgian e-ID card, the data controllers' security obligations according to data protection law, the roles and liabilities of the users of user-centric identity management applications, and the regulation surrounding the use and offering of anonymous communication services.

# List of Contributions

| | |
|---|---|
| Introduction | COSIC, ICRI, DistriNet and McGill |
| Digital credentials | McGill, DistriNet and COSIC |
| Evaluation of Anonymity Systems | COSIC |
| Privacy Enhanced Applications | COSIC, DistriNet and ICRI |
| Data storage | COSIC |
| Legal aspects | ICRI and COSIC |
| Conclusions | COSIC, ICRI, DistriNet and McGill |
| | |
| Editor | Claudia Diaz (COSIC) |

# List of ADAPID Publications covered in this report


1. Mira Belenkiy, Melissa Chase, Markulf Kohlweiss, and Anna Lysyanskaya. *P-signatures and Noninteractive Anonymous Credentials.* In Proceedings of the 5th Theory of Cryptography Conference (TCC'08), R. Canetti (ed.), Springer LNCS 4948, pp. 356-374, 2008.

2. Jan Camenisch, Markulf Kohlweiss, Alfredo Rial and Caroline Sheedy. *Blind and Anonymous Identity-Based Encryption and Authorised Private Searches on Public-key Encrypted Data.*, accepted to the 12th International Conference on Practice and Theory in Public Key Cryptography (PKC 2009).

3. Bart De Decker, Mohamed Layouni, Hans Vangheluwe, and Kristof Verslype, *A Privacy-preserving eHealth Protocol Compliant with the Belgian Healthcare System*, EuroPKI, Lecture Notes in Computer Science, vol. 5057, Springer, 2008, pp. 118–133.

4. George Danezis, Claudia Diaz, Sebastian Faust, Emilia Käsper, Carmela Troncoso and Bart Preneel. *Efficient Negative Databases from Cryptographic Hash Functions.* In Proceedings of the 10th Information Security Conference (ISC'07), J. Garay et al. (Eds.), Springer LNCS 4779, pp. 423-436, 2007.

5. Claudia Diaz, Carmela Troncoso, and George Danezis. *Does additional information always reduce anonymity?.* In Workshop on Privacy in the Electronic Society 2007, T. Yu (ed.), ACM, pp. 72-75, 2007.

6. Claudia Diaz, Carmela Troncoso, and Andrei Serjantov. *On the Impact of Social Network Profiling on Anonymity.* In Privacy Enhancing Technologies Symposium, Lecture Notes in Computer Science 5134, N. Borisov, and I. Goldberg (eds.), Springer-Verlag, pp. 44-62, 2008.

7. Claudia Diaz, Eleni Kosta, Hannelore Dekeyser, Markulf Kohlweiss, and Girma Nigusse. *Privacy preserving electronic petitions.* Submit-

ted to the journal of Identity in the Information Society, 14 pages, 2008.

8. Claudia Diaz, and Pim Tuyls. *Privacy-enhanced biometrics for e-ID cards.* COSIC technical report, 10 pages, 2007.

9. Claudia Diaz, Carmela Troncoso and Bart Preneel. *A Framework for the Analysis of Mix-Based Steganographic File Systems.* In 13th European Symposium on Research in Computer Security (ESORICS 2008), S. Jajodia, and J. Lopez (Eds), Springer LNCS 5283, pp. 428-445, 2008.

10. Benedikt Gierlichs, Carmela Troncoso, Claudia Diaz, Bart Preneel, and Ingrid Verbauwhede. *Revisiting A Combinatorial Approach Toward Measuring Anonymity.* In Workshop on Privacy in the Electronic Society 2008 ACM, ACM, 5 pages, 2008.

11. Markulf Kohlweiss and Alfredo Rial. *Universally Composable Adaptive Priced Oblivious Transfer.* COSIC technical report, 2009.

12. Mohamed Layouni, *Accredited Symmetrically Private Information Retrieval.* IWSEC'07, Lecture Notes in Computer Science, vol. 4752, Springer, 2007, pp. 262–277.

13. Mohamed Layouni, Maki Yoshida, and Shingo Okamura, *Efficient Multi-Authorizer Accredited Symmetrically Private Information Retrieval*, ICICS'08, Lecture Notes in Computer Science, vol. 5308, Springer, 2008, pp. 387–402.

14. Carmela Troncoso, Benedikt Gierlichs, Bart Preneel, and Ingrid Verbauwhede. *Perfect Matching Disclosure Attacks.* In Privacy Enhancing Technologies Symposium, Lecture Notes in Computer Science 5134, N. Borisov, and I. Goldberg (eds.), Springer-Verlag, pp. 2-23, 2008.

15. Carmela Troncoso, George Danezis, Eleni Kosta, and Bart Preneel. *PriPAYD: Privacy Friendly Pay-As-You-Drive Insurance.* In Workshop on Privacy in the Electronic Society 2007, T. Yu (ed.), ACM, pp. 99-107, 2007.

16. Carmela Troncoso, Danny De Cock, and Bart Preneel. *Improving Secure Long-Term Archival of Digitally Signed Documents.* In Proceedings of the 4th International Workshop on Storage Security and Survivability (StorageSS 2008) ACM (in print), 10 pages, 2008.

17. Brendan Van Alsenoy and Danny De Cock, *Due processing of personal data in eGovernment? A Case Study of the Belgian electronic identity card.* Datenschutz und Datensicherheit, March 2008, 178-183.

18. Kristof Verslype and Bart De Decker. *Ubiquitous Privacy-Preserving Identity Managment.* Proceedings of The Ifip Tc 11 23rd International Information Security Conference. Milan (Italy), Sep 8-10, 2008. pp. 589-603.

19. Kristof Verslype, Bart De Decker, Vincent Naessens, Girma Nigusse, Jorn Lapon and Pieter Verhaeghe. *A Privacy-Preserving Ticketing System.* Data and Applications Security XXII, 22nd Annual IFIP WG 11.3 Working Conference on Data and Applications Security, London, UK, July 13-16, 2008. LNCS, vol. 5094, pp. 97-112. Springer, Berlin (2008)

20. Kristof Verslype, Jorn Lapon, Pieter Verhaeghe, Vincent Naessens, Bart De Decker. PetAnon: A Privacy-Preserving e-Petition System Based on Idemix. Report CW 522, Leuven, Belgium, October, 2008

21. Pieter Verhaeghe, Jorn Lapon, Bart De Decker, Vincent Naessens and Kristof Verslype. *Security and Privacy Improvements for the Belgian eID Technology.* Proceedings of The Ifip Tc 11 24th International Information Security Conference. Pafos, (Cyprus), May 18-20, 2009.

# Contents

# Chapter 1

# Introduction

The goal of the ADAPID project is to study and develop advanced applications for e-ID cards, focussing mainly on privacy and security aspects of data storage, e-government and e-health applications. To enable secure and privacy preserving applications, we have investigated a number of basic technologies as well as the legal aspects of using the e-ID for providing services. This report presents a summary of the basic research done within the project in the second and third years, and reports on twenty technical reports and scientific publications by the project partners. A wide range of research problems have been tackled: identity management technologies based on digital credentials; models for measuring privacy; privacy preserving communication infrastructures; and secure storage and retrieval of information. Additionally, we propose several privacy enhanced applications that use the e-ID as source of authentication, and are built with some of the studied technologies. The legal aspects relevant to applications using the e-ID for authentication have also been explored. The rest of this section motivates and summarizes the main research results that can be found in the report.

Digital credentials are an essential building block in the design of privacy-preserving applications. They enable powerful, flexible and user-centric identity management with privacy enhanced properties that previous technologies, such as PKI or paper-based credentials, could not offer. Our research in this area focuses on developing new techniques that offer users more flexibility and control over their data. For instance, we have designed protocols that allow users to control access to their remotely-stored data according to a self-defined privacy policy, without the storage server learning the access pattern to their data. These protocols could be useful in a variety of application domains and we have shown as an example how such protocols can be useful in e-health systems. More details are given in Sections 2.2, 2.3, and 2.4. Our second line of research, which we describe in Section 2.5, is focused on the design of more efficient and generic credential systems.

Anonymous communications is a young field of research that aims to

protect the communication layer from malicious eavesdroppers who perform traffic analysis. Protecting the communication layer is a requirement when building privacy enhanced applications, as otherwise the information concealed using cryptography or other privacy techniques could be uncovered through the analysis of communication data. Our research in this area is mainly directed to the evaluation of anonymity systems. We present in Chapter 3 two papers in which we study the amount of information that an adversary observing the network can extract when she has access to context information about the participants in the communications. This information can be extracted from social networks (like Facebook, LinkedIn, etc), the content of the messages, or any other means available to the attacker. Besides, we revisit previous work on how to apply combinatorial methods to measure the degree of anonymity given by a system. Finally, we develop an attack on a specific anonymity system based on mix networks, which demonstrates that considering all participants in the system at once gives a considerable advantage to an attacker with respect to considering them individually.

Secure data storage is a key building block in the construction of e-health, e-government and e-commerce applications, as they all require the storage and access to confidential information. Moreover, the progress of information technologies has lead to a situation where data collection and data processing can be automated, which facilitates the creation of personal profiles that relate the information that users disclose when carrying out different online transactions. Consequently, data storage mechanisms should provide not only conventional properties like confidentiality, integrity or authentication, but also methods that, without overlooking efficiency, permit to store, search and retrieve data in a privacy-preserving way. We address the construction of mechanisms that protect from different types of adversarial behavior. In Section 4.2 we present a Secure Long-Term Archival System (SLTAS). SLTASs ensure eternal integrity and authentication by preserving the validity of today's digital signatures in a distant future. Section 4.3 introduces two designs of negative databases, where it is hard to list all the existing entries while still being possible to efficiently test if an entry is present. In Section 4.4 we propose a Steganographic File System (SFS), which is intended to hide files from an adversary that first monitors the file storage and then coerces the file owner. We describe a public key encrypted database that admits authorized private searches and private data retrieval in Section 4.5, and finally we present a Priced Oblivious Transfer scheme, which allows a buyer to acquire digital goods without disclosing to the vendor which particular item is being retrieved.

Part of the ADAPID research work has been devoted to the design of applications based on a combination of the Belgian e-ID and the technologies studied in the project. Our results suggest that although challenging, the design of secure and privacy enhanced applications for the e-ID is possible

if the adequate techniques are combined in an appropriate way. This has been further demonstrated by the proof-of-concept implementations of some of our application designs, which have been developed within the project in the context of the applications work packages. Chapter 5 presents our theoretical research on applications, and includes: two schemes built around the Belgian e-ID that implement secure and privacy enhanced electronic petitions; a privacy friendly Pay-As-You-Drive insurance system that prevents location tracking; an online credential vault for the secure storage of credentials; and two privacy-preserving ticketing systems. Additionally, we present a review of the privacy and security issues of including biometric templates in e-ID cards and study the possibility of using privacy-enhanced biometrics techniques.

From the legal perspective, we have investigated three topics, which are described in Chapter 6. In first instance, the privacy risks presented by the Belgian e-ID card are discussed in light of the controller's security obligation under data protection law. The second item of legal research relates to the current state of the regulation surrounding the use and offering of anonymous communication services. Finally, consideration is given to emerging user-centric identity management applications in web 2.0 and the roles and liabilities of their users.

# Chapter 2

# Privacy-Preserving Credentials and Applications

## 2.1 Introduction

Digital credentials [Cha85a, CP92, Bra94, Bra00, CL01a, CL04a, CHK+06a] are an essential ingredient in today's communication infrastructure. They allow users to interact in a secure and authenticated way with a variety of players including government, merchants, hospitals, friends etc. There are two main families of credentials in the literature: traditional public key credentials, and the more recent privacy-preserving credentials.

The X.509 public key certificate standard [IT05] is one of the best examples of traditional public key credentials. The X.509 public key infrastructure – one of the most widely used PKI standards nowadays, and the same standard being used in the current Belgian e-ID card – credentials are issued to users in the form of identity certificates. These certificates contain various information about the identity of the credential holder, and in particular a unique *serial number*. Although useful for the efficient indexing of data, this serial number could be potentially harmful for the privacy of the credential holder, since it allows any party to monitor and link all of his (the credential holder) activities across multiple application domains. Another shortcoming of traditional digital credentials, such as X.509 PKI certificates, is the fact that they are easy to clone and copy, and using them without proper safeguards could lead to serious security problems.

To address the above shortcomings, and reconcile privacy with security requirements, a new class of identity management tools has been built. These tools are based on the so-called privacy-preserving digital credentials [Cha85a, CP92, Bra94, Bra00, CL01a, CL04a]. In a privacy-preserving digital credential system, one can generally distinguish three types of players: a certification authority, a user, and a verifier. In some cases, the certification

authority and the verifier are controlled by the same entity. The certification authority issues a credential to a user who fulfills certain conditions. In exchange for goods and services, the user may be required to prove, to a verifier (service provider), possession of a valid credential from the certification authority. The user may also be required to prove a predicate on the attributes encoded in his credential. The service provider may later decide to deposit a transcript of the interaction it had with the user, to the certification authority.

Credential systems are very heterogeneous, and may have very special requirements specific to the application at stake. Nevertheless, it can be conjectured that the following set of requirements are commonly desired in most credential systems, and can therefore be used as a basis for research and development of generic solutions:

1. Non-forgeability: It should not be possible to forge, with non-negligible effort, a credential on behalf of a certification authority, or to alter the attributes already encoded by the CA in a previously issued credential.

2. Security: It should not be possible for a credential holder to prove, with non-negligible effort, a property or a predicate that is, in reality, not satisfied by the attributes encoded in his CA-issued credential,

3. Non-appropriation: It should not be possible for an individual to use, with non-negligible effort, a credential that does not belong to him. This is to prevent unlawful impersonations and identity thefts.

4. Non-transferability: There should be mechanisms in place to discourage credential holders from sharing their credentials with third parties.

5. Non-modifiability: Any modification to a credential showing transcript should be detectable with overwhelming probability. This property is desirable to preserve evidence, and prevent framing.

6. Privacy with respect to the CA: The certification authority should not be able to link the showing transcripts of a credential, to the issuing protocol instance that generated it.

7. Privacy with respect to the Verifier: A verifier should not be able to learn any information about the attributes embedded in the credential being shown, beyond what the credential holder willfully reveals and the apriori knowledge.

8. Selective disclosure: a credential holder should be able to selectively disclose any partial information or property about the identity attributes embedded in his or her credential, without necessarily revealing their exact values.

9. Selective depositing: It is desirable sometimes for verifiers to be able to deposit showing transcripts that reveal only partial information about the transaction that took place between the credential holder and verifier. The deposited information should be consistent with the initial showing transcript however.

10. Suitability for smartcard implementations: Special attention should be given to efficiency when designing credential systems for smartcards, because of their limited computation and storage resources.

11. Revocability: It should be possible to revoke credentials in case they are used for abusive behaviour. In some cases, this also means the unveiling of the identity of the credential holder.

12. Unlinkability: It should not be possible for a verifier to link different credential showings by the same credential holder.

The latter requirement can be refined even further, by adding constraints on the number of times a credential can be used before unlinkability is broken. Based on this last criterion, we can distinguish two types of credentials:

1. Multiple-show credentials: they can be shown an infinite amount of times without the showings being linked to each other, or to the issuing protocol instance where they were generated.

2. Limited- or $k$-show credentials, for ($k \geq 1$): they can be shown anonymously up to $k$ times, before the showings being linked to each others. In some contexts, it might be desirable to unveil the identity of the credential holder if the credential is used beyond a predefined number of times.

The area of privacy-preserving credentials continues to evolve in a very rapid way, and we are already starting to see a promising host of concrete implementations [Ide, Hig, MSC08]. In the following sections we provide a brief summary of the new results we have obtained in this area.

## 2.2 Accredited Symmetrically Private Information Retrieval (ASPIR)

- Mohamed Layouni, *Accredited Symmetrically Private Information Retrieval.* IWSEC'07, Lecture Notes in Computer Science, vol. 4752, Springer, 2007, pp. 262–277.

## Summary

Privacy preserving credentials is a technology that allows users to selectively disclose information about data lying under their control. In this Chapter, we consider the problem of protecting access to personal information stored on a remote database lying outside the user's control. Our goal is to control access to this information according to privacy policies defined by the owners of the data. More specifically, we propose a new primitive allowing users to authorize access to their remotely-stored data according to a self-chosen access policy, without the storage server learning information about the access pattern, or even the index of the records being retrieved. The proposed solution, which we call *Accredited Symmetrically Private Information Retrieval* (ASPIR), combines symmetrically private information retrieval and privacy-preserving digital credentials. We present three constructions based on the discrete logarithm and RSA problems.

## Background

In a symmetrically private information retrieval (SPIR) system, there are generally two players: a Sender and a Receiver. The Sender has a database DB of records, and the Receiver submits a query $Q$ to the Sender in order to retrieve a particular record. The main requirement in a SPIR system is privacy for both the Sender and the Receiver. That is, on the one hand the Sender should not learn any information about the index of the record the Receiver is interested in, and on the other hand, the Receiver should not learn any information about the database, beyond the content of the record defined in the query $Q$, and what is already publicly known. In particular, the Receiver should not be able to learn information about more than one record per query. For instance, the Receiver should not be able to learn, through one query, the value of any function on a set of more than one record. SPIR systems have many real-world applications; for instance, consider a scenario where the inventor of a new drug needs information on a number of chemical components that will constitute his final product. This information can be accessed for a fee at some central database. This database could be managed, however, by parties with possibly competitive interests, and the inventor might be concerned that his intellectual property (IP) could be compromised. It is therefore natural, that the inventor might want the content of his queries to remain concealed from the database manager. The latter, on the other hand, wants to be paid for all information retrieved from his database. It is clear that the SPIR system described above, can be a solution to this pair of conflicting requirements.

There are similar applications however, that are closely related to the example above, which cannot be solved by a SPIR primitive. Consider for

example the following e-health scenario where three types of participants are involved: (1) a patient, (2) a medical database containing the health records of patients, and (3) a doctor querying the medical database on patients' health records. The medical database and the doctor can be thought of as the Sender and Receiver, respectively, in a traditional SPIR setting. The requirements in the e-health application are as follows:

1. **Privacy for the Receiver**: The Receiver (doctor) wants to retrieve records from the medical database, without the Sender (DB) learning the index of those records, and thus the identity of his patient.

2. **Privacy for the Sender**: The Sender (DB) wants to be sure that, for each query, the Receiver (doctor) learns information only on one record (defined in the query) and nothing about the other records.

3. **Privacy for the data subject**: In order to comply with privacy legislation, the Sender wants to be sure that the Receiver has a valid reading authorization from the owner of the targeted record (i.e., the patient). We call the latter, *an Authorizer*. Notice that the Sender should not be able to learn the Authorizer's identity, otherwise the first requirement will be violated.

### Contributions

Let $Q$ denote the query the Receiver party submits to the Sender, and let $R$ the Sender's response. The Receiver recovers the answer to his query from $R$. Our main goal here is to make sure that before processing the query $Q$, the Sender is convinced that the Receiver has obtained an explicit consent from the owner of the record defined in $Q$, *without* revealing the identity of this owner (i.e., the Authorizer). The solution we propose, combines three cryptographic primitives: privacy-preserving digital credentials, homomorphic encryption, and SPIR systems. We assume that we have a credential system such as those in [Bra00, CL01a, CL04a], where users are able to show their credentials unlinkably, and to selectively disclose information about their attributes.

More specifically, we assume that the Authorizer has a *CA*-issued identity credential *Cred* containing a set of attributes (ID,Age,$\cdots$). The idea is to first make the Authorizer and Receiver jointly compute the query $Q$, and then have the Authorizer produce a signed proof of knowledge of the secret attributes embedded in *Cred*. Also included in the proof, is a predicate stating that the value of the ID attribute embedded in *Cred* is the same as the one encoded in the query $Q$. The Receiver then deposits the signed proof along with the query to the Sender. The Sender first checks the validity of the proof. If accepted it carries on with the SPIR protocol and processes the query, otherwise it rejects.

As mentioned earlier, the signed proof does not reveal any information about the credential holder, and yet guarantees that the content of the query is consistent with the secret identity attribute embedded in the credential. Furthermore, owing to the fact that it is hard for a computationally bounded adversary to forge credentials, or make proofs about credentials he does not own, the Sender can be sure that the Receiver has indeed obtained an explicit consent from the targeted record's owner.

We present three constructions to solve the accredited SPIR problem. The first is based on a modified version of one of Brands DL-based credentials [Bra00, Section 4.5.2], the ElGamal cryptosystem, and a SPIR system proposed by Lipmaa in [Lip05]. The two additional constructions are variants of the first, and use an RSA-based version of Brands credentials [Bra00, Section 4.2.2], in combination with the ElGamal, and the Okamoto-Uchiyama [OU98] cryptosystems, respectively.

## 2.3 Efficient Multi-Authorizer ASPIR

- Mohamed Layouni, Maki Yoshida, and Shingo Okamura, *Efficient Multi-Authorizer Accredited Symmetrically Private Information Retrieval*, ICICS'08, Lecture Notes in Computer Science, vol. 5308, Springer, 2008, pp. 387–402.

### Summary

In the previous section, we discussed a technique allowing users to authorize access to their remotely stored data, according to a self-defined privacy policy, without the database manager learning the access pattern to their records. The work we present in this section, generalizes the above technique to a setting where each record on the database is co-owned by a number of parties instead of a single one. The protocol we propose is such that the storage server answers a query only if convinced that the Receiver party holds a valid authorization from the owners of the target record. This is achieved without the storage server learning any information about the identity of the target record. We provide constructions that allow a Receiver to retrieve a DB record only if he has authorizations from all owners of the target record (respectively, from a subset of the owners of size greater than a threshold.) We also provide a construction where owners of the same record do not have equal ownership rights, and the record in question is retrieved using a set of authorizations consistent with a general access structure. The proposed constructions are efficient and use a pairing-based signature scheme. The presented protocol is proved secure under the Bilinear Diffie-Hellman assumption.

## Background

The constructions in the previous section, covers a setting where each record in the database is owned by a single user. In many applications, data records are the property of several parties simultaneously rather than a single one. For example, in the healthcare domain, a medical procedure is performed by a *doctor* on a *patient* within the premises of a *hospital*. It may be natural in some jurisdictions that all three parties, namely the patient, the doctor, and the hospital, have a right to the database record documenting the medical procedure. As a result, a Receiver (e.g., a second doctor) who wants to have access to the above record, needs an authorization from all three record owners. With the obtained authorizations, the Receiver should be able to retrieve the target record subject to the following conditions: (1) the Receiver can retrieve the record in question only if he has the approval of all record owners, (2) the Sender is convinced that the Receiver's query is approved by the owners of the target data, without learning any information about the index of the target data, or the identity of the authorizers, and (3) the Receiver cannot retrieve information about records other than the one defined in the submitted query.

The ASPIR schemes of the previous section rely on privacy-preserving digital credentials [Bra00] to protect the anonymity of the authorizer with respect to the Sender. The digital credential primitive has been used in addition to hide the index of the retrieved record, and to guarantee the unforgeability of the issued authorizations. While highly versatile, the digital credentials of [Bra00] do require a certain amount of computations from the different participants, especially the authorizers. Moreover, the ASPIR construction assumes that each record owner possesses a digital credential of the type in [Bra00], and that he is willing to use it to issue authorizations.

## Contributions

In this work, we extend the ASPIR protocol to a context where each database record can have multiple owners. The protocol we present in this chapter has a neater and more generic design, and uses SPIR primitives in a black-box fashion, unlike the construction of the previous section which works specifically for Lipmaa's SPIR scheme [Lip05]. The construction we present here is more efficient than the previous one, and uses a lightweight pairing-based signature scheme similar to that in [BLS01] instead of digital credentials. In this chapter, we also propose a $t$-out-of-$n$ threshold multi-authorizer ASPIR variant, where records can be privately retrieved by a Receiver as long as he has authorizations from $t$ out of the $n$ owners of the target record.

We finally treat a setting where the owners' rights to a record are not necessarily equal. For example one could imagine a setting where an authorization from the patient is sufficient to access his medical record, while

authorizations from *both* the doctor and hospital are necessary to access the same record. The latter could be useful in cases of emergency where the patient is unable to grant an authorization.

## 2.4 Privacy-Preserving eHealth

- Bart De Decker, Mohamed Layouni, Hans Vangheluwe, and Kristof Verslype, *A Privacy-preserving eHealth Protocol Compliant with the Belgian Healthcare System*, EuroPKI, Lecture Notes in Computer Science, vol. 5057, Springer, 2008, pp. 118–133.

### Summary

In previous sections, we have described a number techniques to help users protect their privacy and exert more control over their data. The presented techniques can be used in settings where the information is under the user's control, as well as settings where the information is stored on a remote server beyond the user's control. In this section we apply some of these techniques to a real world application: healthcare. Real world healthcare systems are generally large and overly complex systems. Designing privacy-friendly protocols for such systems is a challenging task. In this section we present a privacy-preserving protocol for the Belgian healthcare system. The proposed protocol protects the privacy of patients throughout the prescription handling process, while complying with most aspects of the current Belgian healthcare practise. The presented protocol relies on standard privacy-preserving credential systems, and verifiable public key encryption, which makes it readily fit for implementation.

### Background

To improve communication in healthcare services and increase the patients' trust, one should design ehealth protocols with both security and privacy in mind. Due to the sensitive nature of health data, such protocols should be based on well established cryptographic primitives, and should provide defences against possible user inadvertencies such as ID card losses.

Designing protocols however, without consideration for the current procedures, practices, and existing infrastructures, represents a real obstacle to the adoption of these protocols, regardless of their ingenuity. This is due in part to the high costs required to change the existing infrastructure before the new system can be used. In some cases the proposed protocols require the elimination of entire parties. Sometimes these parties represent on the ground a government agency or a ministry, and removing them is simply unrealistic.

### Contributions

In this work, we design a protocol that protects the privacy of patients throughout the prescription handling process, while complying with most aspects of the current Belgian healthcare practice [1]. The Belgian healthcare system is a large and complex system with many players who do not necessarily share the same interests. The ehealth protocol we propose protects (1) the privacy of patients by eliminating any information leak that may harm their interests, and (2) the privacy of doctors, their prescription habits, and their interactions with patients.

Moreover, our protocol has mechanisms to handle disputes and retrace fraudsters, all without changing the structure of the current Belgian healthcare practice. Healthcare systems with a structure similar to that of the Belgian system, can benefit from the protocol proposed here modulo a few minor adaptations.

## 2.5 P-signatures and Noninteractive Anonymous Credentials

- Mira Belenkiy, Melissa Chase, Markulf Kohlweiss, and Anna Lysyanskaya. *P-signatures and Noninteractive Anonymous Credentials.* In Proceedings of the 5th Theory of Cryptography Conference (TCC'08), R. Canetti (ed.), Springer LNCS 4948, pp. 356-374, 2008.

### Summary

This paper introduces P-signatures. A P-signature scheme consists of a signature scheme, a commitment scheme, and (1) an interactive protocol for obtaining a signature on a committed value; (2) a *non-interactive* proof system for proving that the contents of a commitment has been signed; and (3) a non-interactive proof system for proving that a pair of commitments are commitments to the same value. The paper gives a definition of security for P-signatures and shows how they can be realized under appropriate assumptions about groups with a bilinear map. It makes extensive use of the powerful suite of non-interactive proof techniques due to Groth and Sahai. P-signatures enable, for the first time, the design of practical non-interactive anonymous credential systems whose security does not rely on the random oracle model. In addition, they may serve as a useful building block for other privacy-preserving authentication mechanisms.

[1] There are auxiliary procedures in the Belgian healthcare system that are not covered here. The proposed protocol can be slightly modified to include them.

## Background

Anonymous credentials [Cha85a, Dam90, Bra99, LRSW99, CL01b, CL02, CL04b] let Alice prove to Bob that Carol has given her a certificate. Anonymity means that Bob and Carol cannot link Alice's request for a certificate to Alice's proof that she possesses a certificate. In addition, if Alice proves possession of a certificate multiple times, these proofs cannot be linked to each other. Anonymous credentials are an example of a privacy-preserving authentication mechanism, which is an important theme in modern cryptographic research. Other examples include popular constructions for electronic cash [CFN90, FY93, CP93, Bra93, EJA$^+$04, CHL05, Wei05, CHL06, CGH06], for group signatures [CvH91, CS97, ACJT00, BBS04b, BW06b, BW07], and for anonymous authentication [TFS04, DDP06, NSN05, TS06, CHK$^+$06b]. In a series of papers, Camenisch and Lysyanskaya [CL01b, CL02, CL04b] identified a key building block usually called "CL-signature" that is frequently used in these constructions. A CL-signature is a signature scheme with a pair of useful protocols.

The first protocol, called *Issue*, lets a user obtain a signature on a committed message without revealing the message. The user wishes to obtain a signature on a value $x$ from a signer with public key $pk$. The user forms a commitment *comm* to value $x$ and gives *comm* to the signer. After running the protocol, the user obtains a signature on $x$, and the signer learns no information about $x$ other than the fact that he has signed the value that the user has committed to.

The second protocol, called *Prove*, is a zero-knowledge proof of knowledge of a signature on a committed value. The prover has a message-signature pair $(x, \sigma_{pk}(x))$. The prover has obtained it by either running the Issue protocol, or by querying the signer on $x$. The prover also has a commitment *comm* to $x$. The verifier only knows *comm*. The prover proves in zero-knowledge that he knows a pair $(x, \sigma)$ and a value *open* such that $\mathsf{VerifySig}(pk, x, \sigma) = \mathsf{accept}$ and $comm = \mathsf{Commit}(x, open)$.

It is clear that using general secure two-party computation [Yao86] and zero-knowledge proofs of knowledge for any NP statement [GMW86], we can construct the Issue and Prove protocols from any signature scheme and commitment scheme. Camenisch and Lysyanskaya's contribution was to construct specially designed signature schemes that, combined with Pedersen [Ped92] and Fujisaki-Okamoto [FO98] commitments, allowed them to construct Issue and Prove protocols that are efficient enough for practical use. In turn, CL-signatures have been implemented and standardized [CVH02, BCC04]. They have also been used as a building block in many other constructions [JS04, EJA$^+$04, CHL05, CHL06, DDP06, CHK$^+$06b, TS06, CGH06, CLM07].

A shortcoming of the CL signature schemes is that the Prove protocol is interactive. Rounds of interaction are a valuable resource. In certain

contexts, proofs need to be verified by third parties who are not present during the interaction. For example, in off-line e-cash, a merchant accepts an e-coin from a buyer and later deposits the e-coin to the bank. The bank must be able to verify that the e-coin is valid.

There are two known techniques for making the CL Prove protocols non-interactive. We can use the Fiat-Shamir heuristic [FS87], which requires the random-oracle model. A series of papers [CGH04a, DNRS03, GK03a] show that proofs of security in the random-oracle model do not imply security. The other option is to use general techniques: [BFM88, DSMP88, BDMP91] show how any statement in NP can be proven in non-interactive zero-knowledge. This option is prohibitively expensive.

This paper presents the first *practical* non-interactive zero-knowledge proof of knowledge of a signature on a committed message. We have two constructions using two different practical signature schemes and a special class of commitments due to Groth and Sahai [GS07]. Our constructions are secure in the common reference string model.

Due to the fact that these protocols are so useful for a variety of applications, it is important to give a careful treatment of the security guarantees they should provide. This paper introduces the concept of P-signatures — signatures with efficient **P**rotocols, and gives a definition of security. The main difference between P-signatures and CL-signatures is that P-signatures have non-interactive proof protocols.

### Contributions

Our main contribution is the formal definition of a P-signature scheme and two efficient constructions.

Anonymous credentials are an immediate consequence of P-signatures (and of CL-signatures [Lys02]). Suppose there is a public-key infrastructure that lets each user register a public key. Alice registers unlinkable pseudonyms $A_B$ and $A_C$ with Bob and Carol. $A_B$ and $A_C$ are commitments to her secret key, and so they are unlinkable by the security properties of the commitment scheme. Suppose Alice wishes to obtain a certificate from Carol and show it to Bob. Alice goes to Carol and identifies herself as the owner of pseudonym $A_C$. They run the P-signature Issue protocol as a result of which Alice gets Carol's signature on her secret key. Now Alice uses the P-signature Prove protocol to construct a non-interactive proof that she has Carol's signature on the opening of $A_B$.

Our techniques may be of independent interest. Typically, a proof of knowledge $\pi$ of a witness $x$ to a statement $s$ implies that there exists an efficient algorithm that can extract a value $x'$ from $\pi$ such that $x'$ satisfies the statement $s$. Our work uses Groth-Sahai non-interactive proofs of knowledge [GS07] from which we can only extract $f(x)$ where $f$ is a one-way function. We formalize the notion of an $f$-extractable proof of knowledge

and develop useful notation for describing $f$-extractable proofs that committed values have certain properties.

# Chapter 3

# Attacks and Evaluation of Anonymity Systems

## 3.1 Introduction

Traditionally, computer security has focused on ensuring the confidentiality, integrity and availability of information. This extends to secure communications, where these properties are mostly achieved through cryptographic means. Protective measures, however, are often applied only to high level protocols, leaving network layer information, such as the identities of the participants in the communication (IP addresses), their location, the amount and timing of data transferred or the duration of the connection, accessible to eventual observers. These, commonly known as traffic data, have long been exploited by traffic analysis techniques to deduce further details about the observed communications. For instance, the browsing habits of a company's employees (e.g. accessing a given patent from a patent database), can be used to infer the company's future lines of investment, thus giving advantage to their competitors. This situation becomes more serious in an e-health context, where privacy of patients can be compromised, or in e-voting, where the ability to link voter and ballot interferes with the very purpose of the application.

Anonymous communications' main goal is the protection of some traffic data, more precisely, aims to hide who speaks to whom. Guaranteeing anonymity in network communications is harder than only achieving a secure channel. There have been proposals in this field for anonymous email [Cha81a] and anonymous Internet browsing [RSG98]. Further research resulted in the deployment of real world systems like Mixminion [DDM03a] for email and JAP [BFK01] or TOR [DMS04] for web browsing, which attract an increasing number of users. Anonymous email networks are based on mixes: routers that hide, through cryptographic transformations and

re-ordering techniques, the relation between input and output messages. However, mix networks do not conceal the identity of the participants in the communication. Traffic analysis techniques take advantage of this situation and exploit the fact that users usually have clear patterns of communication, i.e. that they communicate recurrently with their contacts, to re-establish the links between the participants in the communication. We present in Sect. 3.4 a novel attack on anonymity networks, the Perfect Matching Disclosure Attack. We demonstrate that a passive attacker can trace messages through the network which high accuracy. Furthermore, this attack shows how and attacker observing an anonymity network can obtain precise sending profiles of the users of this network, even when she has no prior information about the users' behavior.

The appearance of attacks, together with the different nature of anonymity systems designs, created the necessity of a metric to measure the performance of the anonymity networks implementations. There has been a lot of work on anonymity systems' evaluation; i.e., how to measure the amount of anonymity they provide to users. Shannon entropy over the distribution of all possible recipients was proposed as metric in [DSCP02, SD02a] and it is the most widely used. However, it is yet not well understood how several sources of information can be combined (e.g., user profiles and traffic data) when using this metric. Part of our research in this area aims at improving this understanding. We present in Sect. 3.2 a paper in which we show how, contrary to the common belief, additional information does not necessarily decrease the uncertainty of an attacker; and in Sect. 3.3 our proposal of using Bayesian Inference as method to compute anonymity when various sources of information are available. Finally, in Sect. 3.5 we revisit combinatorial approaches to measuring anonymity and we point out that considering the identity of senders and receivers may reduce the anonymity provided by the system with respect to the case where messages are considered as coming from independent sources.

## 3.2 Does additional information always reduce anonymity?

- C. Diaz, C. Troncoso, and G. Danezis. *Does additional information always reduce anonymity?.* In Workshop on Privacy in the Electronic Society 2007, T. Yu (ed.), ACM, pp. 72-75, 2007.

### Summary

We discuss information-theoretic anonymity metrics, that use entropy over the distribution of all possible recipients to quantify anonymity. We identify

a common misconception: the entropy of the distribution describing the potential receivers does not *always* decrease given more information. We show the relation of these a-posteriori distributions with the Shannon conditional entropy, which is an average over all possible observations.

## Background

The most widely accepted definition of anonymity was given by Pfitzmann and Hansen in [PH01]: "anonymity is the state of being not identifiable within a set of subjects, the *anonymity set*." The *anonymity set* is "the set of all possible subjects who might cause an action." In other words, subjects are more anonymous as they can hide in a larger crowd.

The adversary of an anonymity system can typically obtain a probability distribution linking an action to all possible subjects who may be related to it. The adversary's uncertainty on the identity of the subject behind an action depends on the number of subjects in the anonymity set, but also on how the probability distribution looks like: as subjects appear more equally likely to be related to the action, the adversary has less information on who might be the real subject linked to it.

The concept of *Shannon entropy* [Sha48] (or simply "entropy") in information theory is a measure of the uncertainty associated with a random variable. Technical measures of anonymity [DSCP02, SD02b] are based on the entropy of the probability distribution linking an action to all possible subjects who may be related to it, and they give a measure of the uncertainty of the attacker. Shannon entropy has been very useful in the evaluation [DP04, DS03] of mix-based [Cha81a] communication. Metrics based on this entropy have also been proposed to measure the anonymity of profiled users [Cla06, DCSP02].

However, some aspects of entropy-based anonymity metrics are not yet well understood, such as the combination of several sources of information. It has been claimed that an adversary with access to more information is *always* able to reduce anonymity [CS06].

## Contributions

In this paper, we show that the combination of user profile information with observations at the communication layer does not *necessarily* lead to a reduction of the attacker's uncertainty. As Shannon proves that $H(Y|X) \leq H(Y)$, the confusion between the adversary's uncertainty (which is represented by the "entropy of a conditional probability distribution") and the "conditional entropy" has led some researchers to mistakenly believe that access to more information in any concrete case must necessarily result in a reduction of the adversary's uncertainty. We explain in this paper the relationship between

the adversary's uncertainty, which is calculated from a particular observation, and Shannon's conditional entropy (which averages all possible cases), and shown an example where the recipient anonymity of a message increases when the adversary combines two sources of information (profiling and network observation).

## 3.3 On the Impact of Social Network Profiling on Anonymity

- C. Diaz, C. Troncoso, and A. Serjantov. *On the Impact of Social Network Profiling on Anonymity.* In Privacy Enhancing Technologies Symposium, Lecture Notes in Computer Science 5134, N. Borisov, and I. Goldberg (eds.), Springer-Verlag, pp. 44-62, 2008.

### Summary

This paper studies anonymity in a setting where individuals who communicate with each other over an anonymous channel are also members of a social network. In this setting the social network graph is known to the attacker. We propose a Bayesian method to combine multiple available sources of information and obtain an overall measure of anonymity. We study the effects of network size and find that in this case anonymity degrades when the network grows. We also consider adversaries with incomplete or erroneous information; characterize their knowledge of the social network by its quantity, quality and depth; and discuss the implications of these properties for anonymity.

### Background

In the last few years defining and quantifying anonymity in the context of communication networks has been a hot research topic. A substantial set of papers focus on the definition of anonymity, others present designs and analysis of new anonymous communication systems or attacks of existing ones. Yet more focus on the theory of mix systems in order to improve our fundamental understanding of anonymity properties which are possible or practically achievable.

This paper belongs to a growing body of work focusing on the anonymity analysis of anonymous communication systems. A substantial part of this literature consists of papers evaluating the effectiveness of mix-based anonymity systems in a theoretical setting; e.g., [Dan04, DS03, SDS02]. Such work often involves assumptions such as "users pick their communication partners uniformly at random" which help with the mathematics of calculating anonymity, and hence aid our understanding and intuition, but do

not necessarily hold in practice. Furthermore, the authors often examine properties of the anonymous communication systems and shy away from incorporating models of users.

This paper takes the fine line between theory and practice and attempts to evaluate the anonymity properties of an abstract anonymous communication system within the practical context of a social network.

### Contributions

The main contribution of this paper is evaluating how the uncertainty in the attacker's knowledge of user profiles affects anonymity. Indeed, we show that arbitrarily small errors in the profiles can lead to arbitrarily large errors in the anonymity probability distribution and hence point to the wrong subjects in the anonymity set. We develop the intuition behind this result and evaluate the errors in the anonymity probability distributions in the context of the social network. We conduct our experiments by simulation which helps us examine realistic scenarios.

## 3.4 Perfect Matching Disclosure Attacks

- C. Troncoso, B. Gierlichs, B. Preneel, and I. Verbauwhede. *Perfect Matching Disclosure Attacks.* In Privacy Enhancing Technologies Symposium, Lecture Notes in Computer Science 5134, N. Borisov, and I. Goldberg (eds.), Springer-Verlag, pp. 2-23, 2008.

### Summary

Traffic analysis is the best known approach to uncover relationships amongst users of anonymous communication systems, such as mix networks. Surprisingly, all previously published techniques require very specific user behavior to break the anonymity provided by mixes. At the same time, it is also well known that none of the considered user models reflects realistic behavior which casts some doubt on previous work with respect to real-life scenarios. We first present a user behavior model that, to the best of our knowledge, is the least restrictive scheme considered so far. Second, we develop the Perfect Matching Disclosure Attack, an efficient attack based on graph theory that operates without any assumption on user behavior. The attack is highly effective when de-anonymizing mixing rounds because it considers all users in a round at once, rather than single users iteratively. Furthermore, the extracted sender-receiver relationships can be used to enhance user profile estimations. We extensively study the effectiveness and efficiency of our attack and previous work when de-anonymizing users communicating through a threshold mix. Empirical results show the advantage of our proposal. We

also show how the attack can be refined and adapted to different scenarios including pool mixes, and how precision can be traded in for speed, which might be desirable in certain cases.

## Background

Traffic analysis exploits traffic data to infer information about observed communications. It is the most powerful known attack against anonymous networks. More precisely, Disclosure (or Intersection) attacks use the fact that users' communication patterns are repetitive to uncover communication relationships between them [AK03a, Dan03].

Previous work on Disclosure Attacks [AK03a, Dan03, DDT07] considers a very simplistic model, where users send messages to a fixed set of contacts through a threshold mix. Users choose amongst their communication partners with uniform probability and the effectiveness of these attacks strongly relies on this model. In this paper we present a new attack, the Perfect Matching Disclosure Attack, that requires no assumption on the users' behavior in order to reveal their relationships. Besides its capability to uncover relations amongst users, *i.e.* their sending profiles, in an arbitrary scenario, we demonstrate the strength of our attack in de-anonymizing individual messages, *i.e.* finding the links between messages arriving to the network and messages leaving it. Our method's advantage stems from the fact that it considers all users in a round at once, rather than single users iteratively. This approach is likely to de-anonymize a large fraction of the set correctly in scenarios where a per user approach fails with high probability.

## Contributions

The main drawback of previously published practical Disclosure Attacks is their susceptibility to changes in the user behavior model. Each of them seems to be optimized for a specific and restricted scenario. Our first contribution is a more general user behavior model, where the number of users' friends and the distribution of sending probabilities toward them is not restricted.

Our second contribution is the Perfect Matching Disclosure Attack, that achieves a high rate of success when tracing messages sent through a threshold mix in arbitrary scenarios. Its accuracy arises from the fact that it considers information about all senders participating in a round simultaneously, rather than focusing on individual users iteratively. We empirically compare it with previous work in terms of effectiveness and show that our proposal yields better results when de-anonymizing the sender of a given message in a generic scenario.

The second advantage of the PMDA over previous work is its enhanced ability to estimate user profiles. Concerning a very restrictive user behavior

model we empirically confirm that the PMDA yields a better separation of friends and non-friends than previous work. With respect to a generic scenario we show that the PMDA reliably identifies users' friends when previously proposed methods fail.

## 3.5 Revisiting A Combinatorial Approach Toward Measuring Anonymity

- B. Gierlichs, C. Troncoso, C. Diaz, B. Preneel, and I. Verbauwhede. *Revisiting A Combinatorial Approach Toward Measuring Anonymity.* In Workshop on Privacy in the Electronic Society 2008 ACM, ACM, 5 pages, 2008.

### Summary

Recently, Edman et al. proposed the *system's anonymity level* [ESY07], a combinatorial approach to measure the amount of additional information needed to reveal the communication pattern in a mix-based anonymous communication system as a whole. The metric is based on the number of possible bijective mappings between the inputs and the outputs of the mix. In this work we show that Edman et al.'s approach fails to capture the anonymity loss caused by subjects sending or receiving more than one message. We generalize the *system's anonymity level* in scenarios where user relations can be modeled as yes/no relations to cases where subjects send and receive an arbitrary number of messages. Further, we describe an algorithm to compute the redefined metric.

### Background

The goal of anonymous communication systems is to hide the correspondence between communication partners, such that an adversary cannot determine who is sending messages to whom. Anonymous communication systems are usually built with mixes [Cha81a, DDM03a, MCPS03a] or onion routers [DMS04, GRS96, RSG98], black boxes whose objective is to hide the correspondence between input and output messages or streams.

The emergence of anonymous communication systems led to the need for anonymity metrics to evaluate and compare different designs. Based on the definition of anonymity proposed by Pfitzmann and Hansen [PH01] "Anonymity is the state of being not identifiable within a set of subjects, the *anonymity set*" information-theoretic metrics were independently proposed by Diaz et al. and Serjantov and Danezis in [DSCP02, SD02b]. These metrics are based on Shannon entropy [Sha48], and express the uncertainty of

an adversary with respect to the sender or recipient of a given message. Several variations of information-theoretic metrics for anonymity have followed: Tóth et al. [THV04] propose using min-entropy and max-entropy for measuring local anonymity; Clauß and Schiffner [CS06] propose to use Rényi entropy [Rél61] as a generalization of Shannon, min- and max-entropy; Deng et al. [DPW06] suggest using relative entropy; and Zhu and Bettati [ZB05] propose an anonymity metric based on mutual information. A different approach was followed by Edman et al. [ESY07]. Instead of computing the size of the (sender or recipient) anonymity set for a given message, they consider simultaneously all incoming and outgoing messages in an anonymous communication system. Combinatorial approaches have also been used to model unlinkability [FMP07] and in the context of disclosure attacks [AK03b].

## Contributions

In this paper we revisit Edman et al.'s *system's anonymity level* and show that it does not capture the anonymity loss caused by subjects sending or receiving multiple messages. We propose a generalization of the metric in scenarios where user relations can be modeled as yes/no relations taking multiple messages per subject into account. We provide a divide and conquer algorithm to compute the redefined *system's anonymity level*. The key difference between our approach and Edman et al.'s is that we consider relationships between senders and recipients, rather than between individual input and output messages.

While our observation also applies to anonymity metrics that measure the size of sender or recipient anonymity sets, we note that it is trivial to adapt these metrics to account for multiple messages per subject. This was not explicitly addressed in some of the first works [DP04, DSCP02, SD02b], but later papers [DTS08] do consider multiple messages per subject.

# Chapter 4

# Secure and Privacy-Enhanced Storage

## 4.1 Introduction

Nowadays, data is created, processed and stored digitally, and thus secure data storage is indispensable for the development of the information society. A secure data storage mechanism must provide confidentiality and integrity in order to ensure that data is neither accessed nor modified by unauthorized parties. Another desirable property is message authentication, which provides assurance on the identity of the generator of data. In the race to achieve these properties, however, we should not overlook availability, i.e., data must be efficiently accessible by authorized entities.

Furthermore, the progress of information technologies has provided efficient tools to collect and share data. When using electronic communication services, users reveal information that can be utilized to describe and individual. Among this information not only personal data like name, address or phone number should be included, but also every kind of data, e.g., the search terms used when browsing the web, that can give hints on their political membership, religion or behavior. This data disclosure, along with data collection tools, leads to a situation where personal profiles are easily created and sold.

From a legal point of view, privacy protection is of major importance [eu-95]. Therefore, since the kind of information queried can be useful to characterize a user, secure data storage mechanisms that allow for privacy-preserving methods for remotely searching and retrieving data are necessary. These privacy-preserving methods must ensure that neither the search terms nor the retrieved data is revealed to the data holder, while still providing confidentiality, integrity, and, when needed, authentication.

We address the construction of secure and privacy-preserving data stor-

age mechanisms for different scenarios and by considering different types of adversarial behavior. In Section 4.2 we present a Secure Long-Term Archival System (SLTAS). SLTASs are intended to ensure eternal integrity and authentication by preserving the validity of today's digital signatures in a distant future. An SLTAS takes into account the fact that today's signing algorithms will show vulnerabilities in the future. Our construction is the first that provides at the same time proof that the signing key was valid when it was created, that the signed document remained unmodified and that it was signed in an indisputable time interval.

We then show two negative database constructions. A negative database is a database in which it is hard to list all the existing entries, while still being possible to efficiently test if an entry is present. This ensures data protection even when an adversary takes full control of the database, and it allows handing the database to third parties interested in performing searches. Our constructions improve previous results in that they reduce the storage size and they rely on more standard security assumptions.

In Section 4.4 we propose a Steganographic File System (SFS). An SFS is intended to hide files from an adversary that first monitors the file storage and then coerces the file holder, using the information obtained to search for files. The files are hidden by providing several security levels that, when coerced, allow the user to reveal files of some levels while keeping others hidden. Our construction is secure against powerful attacks that break previous solutions.

We propose a public key encrypted database that admits authorized private searches in Section 4.5. It is based on a new concept we call Public Key Encryption with Oblivious Keyword Search (PEOKS), which extends Public Key Encryption with Keyword Search (PEKS). Since the database is public key encrypted, an adversary that compromises it cannot access the data. The secret key is stored in a security server, which is only involved in providing search trapdoors to users that have authorization to search for data described by a keyword of her choice. Our solution improves those based in PEKS in that we propose a mechanism to hide the search results from the database holder and we allow for different authorization policies, e.g., the searcher can prove in zero-knowledge statements about the keyword without revealing it, whereas in previous schemes the searcher has to send the keyword to the security server, which is also in charge of giving authorization.

Finally, we present a Priced Oblivious Transfer (POT) scheme. POT is a primitive intended to provide privacy during purchase in e-commerce applications. A buyer obtains digital goods from a vendor without revealing to the vendor which items are bought, while guaranteeing to the vendor that she pays the right price for each item and that she does not learn any information about other items. Our construction is secure under concurrent composition, whereas previous solutions fail even under sequential composi-

tion. Moreover, it allows the vendor to charge a different price for the same item to different buyers, which can be used to make personalized discounts.

## 4.2 Secure Long-Term Archival

- Carmela Troncoso, Danny De Cock, and Bart Preneel. *Improving Secure Long-Term Archival of Digitally Signed Documents.* In Proceedings of the 4th International Workshop on Storage Security and Survivability (StorageSS 2008) ACM (in print), 10 pages, 2008.

### Summary

Long-term archival of signed documents presents specific challenges that do not need to be considered in short-term storage systems. We present a Secure Long-Term Archival System (SLTAS) that protects, in a verifiable way, the validity of today's digital signatures in a distant future. Moreover, our protocol is the first proposal that provides a proof of when a signature was created, without the possibility of backdating. We include a description of our scheme and an evaluation of its performance in terms of computing time and storage space. Finally, we discuss how to extend our system to achieve additional security properties.

### Background

Nowadays many documents are created as (or transformed into) digital records and plenty of electronic services (e-government, e-commerce, stock exchange, etc.) depend on them. This requires digital repositories where this information can be stored and accessed in a reliable and secure way [McA07]. Traditionally, archival systems aim at ensuring integrity and availability, but even if an archive initially provides these properties, in the course of time the integrity will certainly degrade.

This work is a step forward on the storage of digitally signed documents and on how to preserve the eternal validity of their digital signatures (and thus provide eternal protection of the integrity of the documents). Secure long-term archival of signed documents suffers from the fragile nature of the private signing keys and the validity status of the corresponding public verification key certificates. The algorithms we use today to compute digital signatures will show vulnerabilities in the long run due to progress in cryptology and the increasing computing power [Moo98]. In order to preserve the validity of signatures, a Secure Long-Term Archival System (SLTAS) must implement a mechanism able to provide proof that the following requirements were checked at a particular time, namely before the document was archived:

- the document was signed between two indisputable moments in the past,
- the content of the signed document has not changed during the storage period, and
- the signature was generated with a signing key that was valid when it was created.

The first requirement guarantees that the signing time cannot be shifted before or after the two indisputable moments in the past, a property which is indispensable in some applications. Consider the scenarios of electronic stock market transactions, or betting systems. In these cases, allowing back-dating of an order's signature may open the door to fraud which can have an important monetary impact.

Meeting the last requirement ensures the eternal validity of the signed document, as the revocation status of the signer's certificate may influence the validity of the signed document at a later stage. Only signed documents of which the signer's certificate was not revoked or suspended at the time of verification remain valid forever and are added to the archive. Let us consider a scenario where Alice and Bob use digital signatures to sign a contract. Although both have valid keys at the moment of signing, a year later Alice loses her key and has to revoke her certificate. In this case, the contract remains only valid if, and only if, there is a proof that both signers' certificates were valid when the contract was signed, thus before Alice revoked her certificate.

Protocols to fulfill the second requirement have been presented in [Bla07, BDJB04, BS05, CW04, HK06], yet, to the best of our knowledge, there is no scheme that satisfies all three. We present the first protocol that simultaneously meets all three properties and at the same time requires a minimum level of trust in the SLTAS.

### Contributions

The contribution consists in a protocol that deals with the weakening of cryptography over time and the need to revalidate signed documents and their certificates. Our scheme permits to prove far in the future the validity of a digital signature in the past, thus providing the *eternal* validity of this signature. Furthermore, it allows binding the time span of the signature generation.

Our protocol improves all previous schemes [Bla07, BS05, CW04, HK06, BDJB04], by bounding the time of existence and validity of the signature between two moments in time, instead of just proving that it existed before the moment of archival. Our approach requires the minimum amount of trust in the participating servers and provides full verifiability of the actions

taken by the SLTAS. We have also shown that the time and space overhead associated with our protocol does not present an impediment to its usability. Finally, our protocol can be integrated with other proposed solutions to take into account migration problems, availability, etc.

## 4.3 Negative Databases

- George Danezis, Claudia Diaz, Sebastian Faust, Emilia Käsper, Carmela Troncoso and Bart Preneel. *Efficient Negative Databases from Cryptographic Hash Functions.* In Proceedings of the 10th Information Security Conference (ISC'07), J. Garay et al. (Eds.), Springer LNCS 4779, pp. 423-436, 2007.

### Summary

A negative database is a privacy-preserving storage system that allows efficiently testing if an entry is present, but makes it hard to enumerate all encoded entries. We improve significantly over previous work presented at ISC 2006 by Esponda *et al.* [EAH$^+$06], by showing constructions for negative databases reducible to the security of well understood primitives, such as cryptographic hash functions or the hardness of the Discrete Logarithm problem. Our constructions require only $\mathcal{O}(m)$ storage in the number $m$ of entries in the database, and linear query time (compared to $\mathcal{O}(l \cdot m)$ storage and $\mathcal{O}(l \cdot m)$ query time, where $l$ is a security parameter). Our claims are supported by both proofs of security and experimental performance measurements.

### Background

In a celebrated series of academic papers [EAFH04, EAH$^+$06], which have also attracted wide public interest [eco06], Fernando Esponda *et al.* introduce the concept of *negative databases* to protect the privacy of stored records. A negative database is a representation of a set of records (the positive database) that allows its holder to test whether particular entries are present in the database, but makes it very hard to efficiently enumerate all entries.

Negative databases have a wide range of applications with the potential to enhance privacy: holders of records cannot easily retrieve all entries, and lost or compromised machines do not therefore lead to large scale privacy compromises. As a result, data holders can also share information without fear that the receiver will be able to extract the full contents of the database. As an example the Transport Security Administration can provide a black list of passengers as a negative database to airline companies, which can

use it to check whether passengers to fly with them are on the list. Yet the companies would not be able to extract the full contents of the database unless they can perform an exhaustive search on it.

The protocols of Esponda *et al.* for building negative databases are limited in many ways, and lead to a database of size $\mathcal{O}(l \cdot m)$ entries – where $m$ is the number of entries in the positive database, and $l$ the size of each entry in bits (with the restriction that $l > 1000$ for security, thus leading to total storage requirement of $\mathcal{O}(l^2 \cdot m)$ bits). Furthermore, the security of their scheme relies on generating and not being able to solve hard instances of the 3-SAT problem, which is a non-standard security assumption in computer security and cryptography (but note that the 3-SAT problem is NP-complete). For a detailed security analysis the reader is referred to the original papers.

### Contributions

In this work we present two constructions that provide exactly the same functionality as the original negative database schemes. Our constructions are computationally efficient for all operations and lead to much more compact "negative" representations. We prove the security of our constructions using standard cryptographic reductions to the security of well understood primitives, such as cryptographic hash functions for our first scheme and the family of Discrete Logarithm assumptions for the second scheme. Experimental results demonstrate that implementing our first scheme is straightforward and can efficiently scale to databases of many megabytes.

## 4.4 Steganographic File Systems

- Claudia Diaz, Carmela Troncoso and Bart Preneel. *A Framework for the Analysis of Mix-Based Steganographic File Systems.* In 13th European Symposium on Research in Computer Security (ESORICS 2008), S. Jajodia, and J. Lopez (Eds), Springer LNCS 5283, pp. 428-445, 2008.

### Summary

The goal of Steganographic File Systems (SFSs) is to protect users from coercion attacks by providing plausible deniability on the existence of hidden files. We consider an adversary who can monitor changes in the file store and use this information to look for hidden files when coercing the user. We outline a high-level SFS architecture that uses a local mix to relocate files in the remote store, and thus prevent known attacks [TDDP07] that rely on

low-entropy relocations. We define probabilistic metrics for unobservability and (plausible) deniability, present an analytical framework to extract evidence of hidden files from the adversary's observation (before and after coercion), and show in an experimental setup how this evidence can be used to reduce deniability. This work is a first step towards understanding and addressing the security requirements of SFSs operating under the considered threat model, of relevance in scenarios such as remote stores managed by semi-trusted parties, or distributed peer-to-peer SFSs.

## Background

Steganographic File Systems (SFSs) were first proposed by Anderson et al. in [ANS98]. The goal of these systems is to conceal not just the content of the files they store, but the very existence of some of those files. Steganography is required to protect users from coercion attacks, where they are forced (e.g., under the threat of violence) to disclose their cryptographic keys to the attacker if the existence of files is known. To protect against these attacks, SFSs typically provide the user with several security levels, each associated with a cryptographic key. In case of coercion, the user provides keys to some security levels (thus revealing some files) without leaking information on the existence of hidden security levels (containing hidden files that are indistinguishable from random data).

Some of the previous SFS proposals [ANS98, MK99] were designed to protect against attackers who obtain a few snapshots of the file store sufficiently spaced in time (e.g., customs inspection performed when entering and leaving a country). However, adversaries who permanently monitor the file store are of practical relevance. For example, the model in [ZPT04] considers a shared multi-user file store where a malicious user or system administrator monitors store accesses. And this threat model is particularly relevant for distributed peer-to-peer SFSs [GL04, HR02], given that any eavesdropper in the vicinity of the user can monitor her connections to other peers (each storing some file blocks), and use the traffic information to obtain evidence of hidden files.

StegFS [ZPT04] is, to the best of our knowledge, the only previous proposal of SFS that implements measures to protect against this adversary model. StegFS avoids simple location access frequency analysis by continuously generating dummy accesses to random locations in the store, and by relocating file blocks every time they are accessed. In spite of these measures, previous work [TDDP07] has shown that the low-entropy block relocation technique used in [ZPT04] enables very powerful traffic analysis attacks capable of uncovering virtually any "hidden" files. In order to counter these traffic analysis attacks, SFSs subject to continuous surveillance require some form of high-entropy block relocation strategy. Such relocation strategy can be achieved using mixes [Cha81b], a well-known mechanism for implement-

ing anonymous email services [DDM03b, MCPS03b]. Besides cryptographically changing the appearance of messages, mixes alter the message flow to prevent traffic analysis attacks based on input and output order, a useful property to introduce uncertainty in the block relocation process. This work develops a framework for analyzing mix-based SFSs, and its purpose is to serve as basis for their design and evaluation. We define probabilistic metrics that characterize the security of an SFS by its unobservability and (plausible) deniability, present methods to analyze whether evidence of hidden files is leaked to the adversary, and validate our analysis through experiments. Our results highlight the power of traffic analysis techniques and the challenge of achieving acceptable levels of security against adversaries who can monitor SFS accesses.

### Contributions

This work studies the security of Steganographic File Systems (SFSs) intended to protect the user against adversaries who monitor accesses to the store. We present an architecture of SFS that uses pool mixes to achieve high-entropy block relocation, and prevent known vulnerabilities to traffic analysis attacks [TDDP07] that exploit low-entropy relocation algorithms [ZPT04].

We define probabilistic metrics to quantify the unobservability and (plausible) deniability provided by SFSs against coercion attacks. Building on existing mix analysis techniques [DP04], we present novel traffic analysis methods to evaluate the security of SFSs subject to continuous observation. In order to validate our approach we implement a MixSFS simulator, examine each step of the attack process, and compute results for unobservability and deniability in an experimental setup. Although we use as example in our analysis a particular type of pool mix, it is trivial to adapt our analysis to other probabilistic relocation mechanisms. The methods introduced here serve as basis for further work on the design and evaluation of traffic analysis resistant SFSs. We note that previous designs have given little or no attention to preventing these types of attacks, in spite of sometimes relying on architectures that use distributed peer-to-peer storage [GL04, HR02], or remote stores observable by third parties, and are thus vulnerable to the adversary and attacks described here.

## 4.5 Authorised Private Searches on Public Key Encrypted Data

- Jan Camenisch, Markulf Kohlweiss, Alfredo Rial and Caroline Sheedy. *Blind and Anonymous Identity-Based Encryption and Authorised Private Searches on Public-key Encrypted Data.*, accepted to the 12th

International Conference on Practice and Theory in Public Key Cryptography (PKC 2009).

## Summary

Searchable encryption schemes provide an important mechanism to protect data while keeping it available to be searched and accessed. In a common approach for their construction, the encrypting entity chooses one or several keywords that describe the content of each encrypted record of data. To perform a search, a user obtains a trapdoor for a keyword of her interest and uses this trapdoor to find all the data described by this keyword.

We present a searchable encryption scheme that allows users to privately search by keywords on encrypted data in a public key setting and decrypt the search results, even if they do not hold the corresponding secret key. To this end, we define and implement two primitives: public key encryption with *oblivious* keyword search (PEOKS) and *committed blind anonymous* identity-based encryption (IBE). PEOKS is an extension of public key encryption with keyword search (PEKS) in which users can obtain trapdoors from the secret key holder without revealing the keywords. Furthermore, we define committed blind trapdoor extraction, which facilitates the definition of authorisation policies to describe which trapdoor a particular user can request. We construct a PEOKS scheme by using our other primitive, which we believe to be the first blind and anonymous IBE scheme. We apply our PEOKS scheme to build a public key encrypted database that permits authorised private searches, i.e., neither the keywords nor the search results are revealed.

## Background

Vast quantities of sensitive personal data are retained for the purpose of network forensics and cyber investigations [eu-06a]. The advantages of the availability of such data for the investigation of serious crimes and the protection of national security are considerable. However, these advantages must be counterpoised by the dangers that such data could fall into the wrong hands.

The encryption of retained data is a desirable counter measure against data theft. But how, then, can the investigator, such as the police or a secret service, search the data without having to decrypt the whole database? What if the investigator should only be given access to data that fulfills certain criteria? This seems to be a hard problem, as the criteria themselves may be sensitive and thus requiring protective measures, such as encryption. Moreover, a secret service is often reluctant to reveal the type of queries it wants to run on the encrypted database.

We consider a scenario in which an investigator searches for data described by multiple keywords without revealing the keywords or the search results to the database server. This scenario is akin to the private searching of streaming data presented in [OI07]. While in [OI07] the data is searched as it is generated (and can thereafter be discarded), in our scenario data is first stored in encrypted form and can be searched at a later stage. To provide a high level of security we make use of asymmetric cryptography. The retaining server only possesses the public encryption key (and cannot decrypt the retained data itself). In this way, data that is already encrypted remains secure even against a strong adversary that breaks into the database server. The decryption key is stored by a security server, which will only be involved when executing search queries.

As the details of queries made are to be obscured even from the security server, it is necessary to impose some restrictions on the investigator. Thus we introduce some checks and balances to avoid abuse by overzealous or malicious agents. One obvious restriction is in the number of queries that the investigator can make. An unreasonable number of requests may be an indication of abuse. Another restriction that we consider is to involve a judge in granting search warrants to the investigator. The keyword is still hidden, but the security server is guaranteed that a judge (or another authority figure) has approved the search for a specific keyword.

In [WBDS04] the authors build an encrypted and searchable audit log. They propose two schemes, one based on symmetric encryption and one based on asymmetric encryption. They conclude that asymmetric encryption provides better security, as it reduces the amount of trust that needs to be placed in the encrypting entity. Our work can be seen as an extension of their asymmetric scheme with the possibility to obliviously search the encrypted database. For the symmetric case, in which the audit log server knows all the information needed for decrypting the database, the problem of performing oblivious searches is covered by [CGN98, OK04]. The problem of oblivious searching on public key encrypted data is more difficult.

The asymmetric searchable encryption scheme in [WBDS04] is based on identity-based encryption (IBE) [BF01]. The keywords themselves are used to encrypt the database, i.e., they are the identity strings of the IBE scheme. The anonymity property of Boneh-Franklin IBE scheme [BF01] ensures that a ciphertext does not leak the identity string used to generate the encryption. The security server holds the master secret key that is used to derive the secret keys corresponding to the keywords that are needed for searching. A similar technique for searchable encryption was formalized as public key encryption with keyword search (PEKS) by [BDCOP04]. In PEKS, the derived keys are referred to as search trapdoors, which can be given to third parties to grant them search rights.

### Contributions

When trying to build an oblivious search mechanism for such a database we have to address two difficulties: hiding the keywords from the security server and hiding the search results from the database. For the former, we present two new cryptographic primitives. The first primitive is *committed blind anonymous* IBE. In this context, *anonymous* means that the ciphertext does not leak the key (identity) under which it was encrypted [Gen06, BW06a] and *blind* means that a user can request the decryption key for a given identity without the key generation entity learning the identity [GH07a]. The work of [GH07a] describes how to construct blind key derivation protocols for [BB04] and [Nac07, CS05], but these schemes are not anonymous. Moreover, it is much harder to derive a blind key derivation protocol for the Boneh-Franklin IBE scheme [BF01] used in [WBDS04], and we are interested in IBE schemes that do not require random oracles for their security proofs. (As shown by [CGH04b, DNRS03, GK03b], a scheme may be insecure even if proven secure in the random oracle model.) As a corollary to our results, we obtain the first instantiation of [WBDS04] secure without random oracles.

We design a committed blind anonymous IBE scheme based on the anonymous IBE scheme due to [BW06a]. As the scheme in [BW06a] is only selective ID secure, we extend it with adapted ID security and prove the modified scheme secure. For the modified scheme we design a blind key extraction protocol. This leads to the first blind anonymous IBE scheme we are aware of. We extend the definition of blind IBE to allow for the derivation of a secret key for a committed identity.

The second primitive we present is public key encryption with oblivious keyword search (PEOKS), which we implement using our committed blind anonymous IBE scheme. First, we extend the definition of PEKS to incorporate the encryption of a secret message when computing a searchable encryption. This secret message can contain a symmetric key, which allows PEKS to be used directly in settings such as [WBDS04]. Then we define blind key extraction with committed keywords, which facilitates the use of a policy that states for which keywords a trapdoor can be extracted while still keeping them hidden from the trapdoor generation entity.

In order to hide the search results from the database one could in theory download the whole database and then use PEOKS to do the search. This is inefficient. We describe a data structure that allows to use private information retrieval (PIR) [CGKS95] to improve the communication efficiency of the search.

## 4.6 Priced Oblivious Transfer

- Markulf Kohlweiss and Alfredo Rial. *Universally Composable Adaptive Priced Oblivious Transfer.* Technical report COSIC, 2009.

### Summary

A $k$-out-of-$N$ Oblivious transfer (OT) scheme is a two-party protocol between a sender with messages $m_1, \ldots, m_N$ and a receiver with selection values $\sigma_1, \ldots, \sigma_k \in \{1, \ldots, N\}$, where the receiver obtains $m_{\sigma_1}, \ldots, m_{\sigma_k}$ without learning anything about other messages, whereas the sender does not learn anything about $\sigma_1, \ldots, \sigma_k$. Priced Oblivious Transfer (POT) is a generalization of OT where receiver pays for the messages without the sender learning the amount paid. Both OT and POT admit an adaptive variant where receiver chooses $\sigma_i$ after receiving $m_{\sigma_{i-1}}$, which enables applications such as privacy-preserving e-commerce.

We present an adaptive OT scheme and we further modify it to construct an adaptive POT scheme. Both constructions are universally composable, optimal in terms of rounds of communication and, after an initialization phase of complexity $O(N)$, both have constant communication and computational cost in each transfer phase. For the construction of the POT scheme we design the first efficient non-interactive proof of knowledge that a value lies in a given interval we are aware of.

### Background

A number of studies [KW99] show that transaction security and privacy concerns are among the main reasons that discourage the use of e-commerce. Although sometimes it is argued that users who claim to be worried about their privacy do not consistently take actions to protect it, recent research [TECA07] demonstrates that, when they are confronted to a prominent display of private information, they not only prefer vendors that better protect their privacy but also are willing to pay higher prices to purchase from more privacy protective websites. Therefore, it is of interest for vendors to deploy e-commerce applications where buyers need to disclose the minimum information needed to carry out their transactions.

So far, the solutions proposed to develop privacy-enhancing e-commerce of digital goods can be roughly divided into two categories: those that hide the identity of the buyer from the vendor (anonymous purchase), and those that hide which goods are bought (oblivious purchase). Anonymous purchase [GA04, LOL04] usually employs anonymous e-cash [Cha82, CHL05, BCKL08a] to construct a system where buyers can withdraw coins from a bank and spend them without revealing their identity. This system has several shortcomings. First, it hinders customer management, e.g., the vendor

cannot easily apply marketing techniques like giving discounts to regular buyers. Second, it does not allow for other methods of payment. Finally, strong anonymity is difficult to achieve and there exist several attacks to reduce it [BFK00].

Oblivious purchase is thus more appealing in scenarios where full anonymity cannot be obtained or when the disadvantages that anonymity causes are important. Oblivious purchase permits effective customer management and it allows for every method of payment. Like anonymous purchase, it has also been shown how to integrate it into existing Digital Rights Management systems [mShWfH]. One can argue that, since the vendor does not know which items he sells, he can find it difficult to know which products are more demanded. However, we note that this information can be obtained from other sources, e.g., by conducting marketing researches.

Oblivious purchase employs the Priced Oblivious Transfer (POT) primitive [AIR01], which is a generalization of the well-known Oblivious Transfer primitive [Rab81] intended to permit private purchases. OT is a two-party protocol between a sender $\mathcal{S}$ and a receiver $\mathcal{R}$, where $\mathcal{S}$ offers a set of messages $m_1, \ldots, m_N$ to $\mathcal{R}$. $\mathcal{R}$ chooses selection values $\sigma_1, \ldots, \sigma_k \in \{1, \ldots, N\}$ and interacts with $\mathcal{S}$ in such a way that $\mathcal{R}$ learns $m_{\sigma_1}, \ldots, m_{\sigma_k}$ and nothing about the other messages, and $\mathcal{S}$ does not learn anything about $\sigma_1, \ldots, \sigma_k$.

POT is a two-party protocol between a vendor $\mathcal{V}$ and a buyer $\mathcal{B}$, where $\mathcal{V}$ sells a set of messages $m_1, \ldots, m_N$ with price $p_1, \ldots, p_N$ to $\mathcal{B}$. Besides the requirements that $\mathcal{V}$ must not learn $\sigma_1, \ldots, \sigma_k$ and $\mathcal{B}$ must not learn anything about other messages, in POT $\mathcal{B}$ must pay prices $p_{\sigma_1}, \ldots, p_{\sigma_k}$ without $\mathcal{V}$ learning anything about the amount of money paid.

Both OT and POT admit an adaptive variant [NP99] ($OT^N_{k\times 1}$,$POT^N_{k\times 1}$), where, in transfer phase $i$, $\mathcal{R}$ or $\mathcal{B}$ may choose $\sigma_i$ after receiving $m_{\sigma_{i-1}}$. The adaptive variant is more suitable for constructing an oblivious database, enabling applications of OT such as treasure hunting, medical record storage or location-based services [NP99], and the deployment of privacy-preserving e-commerce.

The universally composable security paradigm presented in [Can01] provides a framework for representing cryptographic protocols and analyzing their security. Protocols that are proven UC-secure are composable and maintain their security even when they are run concurrently with an unbounded number of arbitrary protocol instances controlled by an adversary.

Traditionally, security in OT was analyzed under a model called half-simulation, where simulation security is required against $\mathcal{R}$, but just stand-alone privacy is required against $\mathcal{S}$. This notion was showed to admit practical attacks against receiver's security in [NP99]. [CNS07], as well as subsequent works [GH07b], present efficient adaptive OT schemes in a full-simulation model. However, these works are not UC-secure because they use black-box simulation with rewinding in their security proofs.

Recently, an adaptive UC-secure OT scheme was proposed in [GH08].

They utilize the approach of assisted decryption used in [CNS07, GH07b], where $\mathcal{S}$ sends to $\mathcal{R}$ a collection of ciphertexts and in each transfer phase helps $\mathcal{R}$ to decrypt one of them. As pointed in [GH08], this approach allows for transfer phases with constant computational and communication cost, and it is suitable to ensure that $\mathcal{S}$ does not change the messages in each transfer phase, which are important properties for constructing an oblivious database. This is in contrast to the approach used in other non-adaptive UC-secure protocols [DNO08, Wag08], where, in each transfer phase, $\mathcal{R}$ hands a set of keys to $\mathcal{S}$, who sends back a collection of ciphertexts such that $\mathcal{R}$ is able to decrypt only one of them.

Despite this recent progress in OT, so far there are not POT schemes whose security is proven within the UC security paradigm or in the full-simulation model. The first POT scheme [AIR01], as well as subsequent works [Tob03], analyze security in the half-simulation model. [DNO08] explains why these protocols fail even under sequential composition and shows a practical attack. Security of both the non-adaptive [IK97, SSR08] and the adaptive [Her08] Generalized Oblivious Transfer (GOT) schemes proposed so far, which can be instantiated as POT schemes, depends on the underlying OT scheme utilized to implement them, but we note that these solutions are rather inefficient.

### Contributions

We present an $OT^N_{k\times 1}$ scheme, and we further modify it to obtain a $POT^N_{k\times 1}$ scheme. As in [GH08], both schemes are UC-secure in a static corruption model under the assumption that there is an honestly generated common reference string, and they do not rely on random oracles in their security proofs. They are optimal in terms of rounds of communication, and each transfer phase has constant computational and communication cost (after an initialization phase of complexity $O(N)$).

Our $OT^N_{k\times 1}$ scheme is based on the one presented in [GH08]. Specifically, we follow their technique of using double trapdoor encryption and we also prove security of ciphertexts under the DLIN [BBS04a] assumption. Nonetheless, unlike [GH08], we make extensive use of P-signatures [BCKL08b], i.e., signature schemes that include efficient non-interactive protocols of signature possession, to let $\mathcal{R}$ prove that she computes her requests honestly. In particular, we employ an slightly modified variant of the multi-block P-signature scheme proposed in [BCKL08a], which is secure under the HSDH and TDH assumptions. (P-signatures also employ Groth-Sahai non-interactive proofs of knowledge [GS08], which rely on either DLIN or SXDH assumptions.) This allows our scheme to have a smaller ciphertext size than the one in [GH08].

We construct a $POT^N_{k\times 1}$ scheme based on our $OT^N_{k\times 1}$, which securely realizes the ideal functionality for POT we define. We follow the approach

in [AIR01] of building a prepaid mechanism where the buyer, after making an initial deposit, computes in each transfer phase a zero-knowledge proof that she updates her account by subtracting the price of the messages she buys, and that the new account is non-negative. For the latter we design a non-interactive range proof of knowledge by applying the efficient interactive range proof recently proposed in [Cha07] to the Groth-Sahai proof system. This is the first efficient non-interactive proof of knowledge to prove that a value lies in a given interval we are aware of.

# Chapter 5

# Privacy Enhanced Applications

## 5.1 Introduction

One of the goals of the ADAPID project is to study and design privacy enhanced applications for the Belgian e-ID card. Most of the work on applications has been done on the applications work packages, which include detailed analysis of the designs as well as the implementation of proof-of-concept demonstrators. This chapter summarizes the more theoretical work done as part of the basic research carried out in the project.

First, we show two alternative designs for the implementation of privacy-enhanced electronic petitions based on the Belgian e-ID card. The first one was developed with in-house code while the second used the Idemix credential toolbox. In both cases, the Belgian e-ID is used for bootstrapping authentication and obtaining a privacy-enhanced digital credential. This digital credential is then used to securely sign electronic petitions. The system design and the properties of the digital credential scheme used ensure that each citizen can only sign a petition once while preserving signer anonymity. Multiple signing of a given petition is detectable, while signatures on different petitions by the same user are anonymous and unlinkable to each other.

Second, we introduce PriPAYD, a privacy friendly Pay-As-You-Drive insurance system. This architecture allows insurance companies to charge personalized premiums dependent on driving habits without compromising the driver's location privacy. The system performs the premium computations at the user side, with the help of a tamper resistant box, so that the insurance company receives aggregated data instead of detailed location data which is sensitive from a privacy point of view.

We continue by presenting an online credential vault which allows the

storage of credentials and other related data of multiple users online, while at the same time maintaining perfect anonymity; the vault server does not obtain any relevant information about its users. This system maximizes the flexibility for the user, as he can use his credentials anywhere at any place. He only needs a smart card. Lost, damage or theft of the smart card does not result in loss of the credentials. These can be retrieved in a privacy-friendly way from the online credential vault.

Then, we present two privacy-friendly methods to buy tickets. The first technique is based on pseudonym certificates, while the second uses the more enhanced Idemix credentials. Deanonimization by a third party (the jurisdiction) is still possible when abuse such as vandalism or participation in riots is committed. The ticket seller still has the flexibility of price discrimination on the basis of personal attributes properties of the user (e.g. age ¡ 18). The ticket seller can also limit the number of tickets a user can buy, while maintaining a maximum level of privacy. Additionally, the access to events by a user who misbehaved can be restricted.

Finally, we present an analysis of privacy and security issues that arise from the inclusion of biometric templates in e-ID cards. We introduce privacy-enhanced biometrics and discuss their advantages in the context of electronic IDs. We discuss the possibility of integrating these technologies with e-ID cards, so that biometric authentication can be a functionality of future generations of the e-ID while guaranteeing the privacy of the citizens as much as possible.

## 5.2 Privacy-Preserving Electronic Petitions

- Claudia Diaz, Eleni Kosta, Hannelore Dekeyser, Markulf Kohlweiss, and Girma Nigusse. *Privacy preserving electronic petitions.* Submitted to the journal of Identity in the Information Society, 14 pages, 2008.

### Summary

We present the design of a secure and privacy preserving e-petition system that we have implemented as a proof-of-concept demonstrator. We use the Belgian e-ID card as source of authentication, and then proceed to issue an anonymous credential that is used to sign petitions. Our system ensures that duplicate signatures are detectable, while preserving the anonymity of petition signers. We analyze the privacy and security requirements of our application, present an overview of its architecture, and discuss the applicability of data protection legislation to our system.

## Background

A petition is a formal request addressed to an authority and signed by numerous individuals. Through petitions, citizens are able to express their support or dissatisfaction with government initiatives, and provide feedback to government institutions. In the physical world, petition signers typically provide a unique identifier (such as the national ID number) together with their handwritten signature, so that fake or duplicate signatures can be eliminated. Given the high cost of collecting and verifying petition signatures by hand, it is not surprising that petitions are increasingly available online. E-petitions present substantial advantages with respect to physical world petitions: it is much easier to reach a large number of people potentially interested on signing them, and the signature verification process can be automated. But they also introduce new security and privacy challenges.

Many of the currently available electronic petitions simply collect the name and national ID number of signers. Given that this information is not secret, it is impossible to check that the petition signer is really providing her own data. In other words, it is not possible to detect cheating, which diminishes the trustworthiness of the petition signature list. To prevent this, some e-petition servers check the IP address of the signer and allow only one signature per IP address. But this disenfranchises legitimate signers who share the IP address with other people (note that in some organizations thousands of users share the same IP address). To ensure that an electronic petition signature is unique and legitimate, it is necessary to use cryptographic means such as digital signatures. Assuming that citizens possess electronic e-ID cards (as is the case in Belgium), an obvious way to implement e-petitions is to have citizens sign them using the key pair available on their e-ID card. However, such a solution is problematic from a privacy point of view. The e-ID public key certificate (needed to verify the digital signature) contains a lot of information about the holder of the card, such as her name, National Registry Number, and date of birth. Revealing all this information for the purposes of signing a petition would definitely be against the data minimization principle, which requires that the processing of data be adequate, relevant and not excessive in relation to the purposes of collection and processing. Additional data protection issues arise when the petitions allow sensitive information to be derived about the user, the processing of which is in general prohibited by data protection legislation. Such information can relate—among other categories of data—to political opinions, religious or philosophical beliefs, all of which are considered as "sensitive personal data" in the European Directive of Data Protection [eu-95].

### Contributions

While signer identification is required in the physical world to ensure the uniqueness and validity of signatures, it is possible to reconcile functional, security and privacy requirements in its electronic version using cryptographic techniques. We propose using the existing PKI-based electronic IDs cards [Coc] in combination with anonymous credential protocols to ensure that (*i*) signatures are legitimate, (*ii*) each citizen can sign a petition at most once, and (*iii*) petition signers are anonymous.

Our e-petition design and implementation uses the Belgian e-ID for initial authentication, and then allows the user to obtain an anonymous credential [Cha85b, CL01c, CHK$^+$06c] that is used to electronically sign petitions on a server. By using anonymous credentials, our demonstrator reconciles two seemingly contradictory requirements: it allows anonymous petition signing, while it imposes restrictions on who is entitled to sign and ensures that each citizen can only sign a particular petition once. Multiple signing of a petition with the same anonymous credential is detectable by our protocols, such that repeated signatures can be eliminated.

## 5.3 Privacy-Preserving Electronic Petitions II

- Kristof Verslype, Jorn Lapon, Pieter Verhaeghe, Vincent Naessens, Bart De Decker. PetAnon: A Privacy-Preserving e-Petition System Based on Idemix. Report CW 522, Leuven, Belgium, October, 2008

### Summary

Electronic petition systems support participation in the democratic decision-making process. However, current petition systems often have serious drawbacks in reliability and anonymity. This paper presents PetAnon, a privacy-preserving petition system that tries to tackle the shortcomings of existing systems. A proof-of-concept implementation is presented that uses the Belgian eID card in a bootstrap procedure, after which users are allowed to sign petitions anonymously, while certain constraints may be imposed.

### Background

In a petition, opinions of people are collected and processed. In paper-based petitions the collection and processing takes a lot of time and effort. Electronic petition systems (e-Petition), however, offer several benefits with respect to the paper-based petitions. e-Petitions enable users to sign petitions anywhere at any time and now reach wider sections of society. Moreover, automatic processing of opinions is faster and less error-prone.

On the other hand, electronic petition systems introduce new problems. Some systems may return unreliable results as for instance a user may sign a petition more than once. Other systems use personal information to prevent multiple signing. However, these systems are not privacy friendly, which is important in petitions where a political, religous, etc. opinion is expressed.

This paper presents PetAnon, a privacy-preserving petition system using Idemix anonymous credentials. PetAnon combines good privacy properties with reliable results.

PetAnon allows petition organizers to provide potential signers with multiple choices. Today, many petitions only have one version: 'in favour'. PetAnon allows to address only a subset of the population (e.g. all women older than 30) and invites the signers to reveal some personal properties such as gender and zip code. Hence, additional statistical information can be collected by the petition organizer. However, this information does not compromise the signer's anonymity. Everyone can verify the correctness of the petition and the petition signer can withdraw or change his vote.

### Contributions

The Idemix based anonymous petition system PetAnon was proposed. The Belgian eID card, that is in casu not privacy friendly was used as a bootstrap. The report showed that it is possible to offer anonymity combined with reliable results for e-Petitions while at the same time, PetAnon is much more versatile than other petition systems. A prototype was developed and revealed that it can be quite demanding w.r.t. computation and storage.

The report showed how unlinkability of a signature with other signatures of that user, as well with the user's identity can be guaranteed, while the user can only vote once. This construction was extended and generalized.

## 5.4 Privacy Friendly Pay-As-You-Drive Insurance

- Carmela Troncoso, George Danezis, Eleni Kosta, and Bart Preneel. *PriPAYD: Privacy Friendly Pay-As-You-Drive Insurance.* In Workshop on Privacy in the Electronic Society 2007, T. Yu (ed.), ACM, pp. 99-107, 2007.

### Summary

Pay-As-You-Drive insurance systems are establishing themselves as the future of car insurance. However, their current implementations entail a serious privacy invasion. This paper presents PriPAYD, a system where the premium calculations are performed locally in the vehicle, and only aggregate

data arrives to the insurance company, without leaking location information. The system is built on top of well understood security techniques that ensure its correct functioning. The paper discusses the viability of PriPAYD in terms of cost, security and ease of certification.

## Background

Pay-As-You-Drive insurance models are hailed as the future of car insurance due to their advantages for users and companies [Lit07, ZB04]. First, the insurance fees applied to each user are fairer than the ones in the pay-by-the-year scheme, as customers are only charged for the actual kilometers they travel. Customers can also reduce their monthly bill by choosing cheap itineraries or by just not using their car. This makes vehicle insurance affordable for lower-income car users (e.g. young people) or for people who wish to have a second vehicle. Second, PAYD policies are socially beneficial, as they encourage responsible driving, decreasing the risk of accidents, which in turn saves money for users and insurers (aside from saving lives). Finally, PAYD has an environmental benefit, as it discourages driving, hence reduces energy consumption and pollution emissions. Due to all these advantages, PAYD insurance policies are supported by motorist associations like the National Motorist Association [Ass98] and the American Automobile Association [Ass]; and they are being widely developed by insurance companies all over the world like Norwich Union [Uni] (UK), Aioi [Aio], Toyota [Cor] (Japan), Hollard Insurance [Ins] (South Africa), etc.

Although PAYD insurance seems to have many advantages, its current implementations involve an inherent threat to user's privacy. The full information used for billing (the time and position where the car was) is gathered by a black box in the car, and transferred to the insurance company (and, in some of the cases, to a third company providing the location infrastructure). In this model, the insurance company has the ability to track any of its users with ease and precision.

## Contributions

The paper proposes PriPAYD, a privacy friendly scheme where the premium computation is done in the car's black box, and only the minimum information necessary to bill the client is received by the insurance company. We provide an overview of our architecture, where well understood techniques are combined to give assurance to the user that the insurance company does not get more information than necessary, while granting him (or a judge in case of dispute) access to all the data. Our techniques also permit easy management and enforcement of the policies by the insurer. A similar case has been pointed by LeMay et al. in [LGGG07] in the field of electric metering, where a remote billing system that preserves user privacy against the

electric company and eavesdroppers is proposed.

## 5.5 Privacy-Preserving Identity Managment

- Kristof Verslype and Bart De Decker. *Ubiquitous Privacy-Preserving Identity Management*. Proceedings of The Ifip Tc 11 23rd International Information Security Conference. Milan (Italy), Sep 8-10, 2008. pp. 589-603.

### Summary

The increasing use of digital credentials undermines the owner's privacy. Anonymous credentials offer a powerful means to improve this. However, more is needed w.r.t. usability. A user will indeed have to manage dozens of credentials in the future: a sporting club credential, a digital driving license, e-tickets, etc. The owner will want to use these credentials wherever and whenever he wants. The credentials must remain manageable as well and, in case of theft or loss, they must become unusable by others and recoverable by the legitimate owner. A possible solution based on smart card or SIM tokens is presented, in which user privacy is maximized. An evaluation reveals both strengths and future challenges.

### Background

A credential is a piece of information attesting to the integrity of certain stated facts: properties about or rights of its owner. Examples are a driving license, money, an identity card and a ticket.

Traditional digital credentials (e.g. X.509 certificates) pose a threat to the privacy of the owner since they generally contain a unique identifier together with other personal data. This data is registered in databases, potentially together with other data (shopping behaviour, medical records, etc.). These data are not only interesting for the entity to which the user shows the credential, but also for insurance companies, for marketeers, etc. Databases containing personal data are thus very valuable and a point of attraction for (internal or external) attackers. They can also get lost, and possibly fall in the wrong hands, as we saw recently in the UK.

Moreover, a user will have to manage dozens of credentials in the future: a sporting club credential, a digital driving license, digital prescriptions, e-tickets for the cinema, etc. The problem of loss of privacy and identity theft will thus aggravate if we do not offer the proper techniques to the users in a practical way. At the same time, the user needs to have access to these credentials wherever and whenever he/she wants. E.g., it is not acceptable that the user has a smart card for each credential or that the credentials

can only be used on a single computer. The credentials must thus remain easily manageable. Credentials must also in case of theft or loss be useless for others and recoverable by the legitimate owner.

Privacy enhancing credentials are being developed and implemented. These allow the user to control the data that will be released. However, more is needed. This paper examines how a user can manage and use credentials such that the above requirements hold. Therefore, a portable user-unique token (e.g. a smart card) is introduced, as well as an online server where credentials can be stored in a privacy-preserving way, while preventing loss and exposure of credentials and related credential data. This paper is the result of an exercise in which we tried to maximize the privacy of the user, while still taking into account the other requirements. The exercise revealed future challenges that must be tackled in order to have a deployable system.

### Contributions

Since authentication and signing is done increasingly electronically, using an increasing number of credentials, two problems arise: 1) the citizen's privacy is at risk and 2) the citizen must be able to easily manage and use his credentials. The first problem can be solved by using anonymous credentials. This paper focuses on the second problem and presents protocols that allow the citizen to manage and use all his credentials using a single smart card. Lost or theft of the smart card does not result in lost or theft of the user's credentials. An analysis is given which shows that in the current context, still certain trust is needed in the device to which the token is connected.

## 5.6 Privacy-Preserving Ticketing Systems

- Kristof Verslype, Bart De Decker, Vincent Naessens, Girma Nigusse, Jorn Lapon and Pieter Verhaeghe. *A Privacy-Preserving Ticketing System.* Data and Applications Security XXII, 22nd Annual IFIP WG 11.3 Working Conference on Data and Applications Security, London, UK, July 13-16, 2008. LNCS, vol. 5094, pp. 97-112. Springer, Berlin (2008)

### Summary

Electronic identity (eID) cards are deployed in an increasing number of countries. These cards often provide digital authentication and digital signature capabilities, but have at the same time serious privacy shortcomings. We can expect that ordering and issuing tickets for events (e.g. soccer matches) will be increasingly done using eID cards, hence, severely threatening the user's privacy. This paper proposes two alternative ticketing systems that

are using the eID card in a bootstrap procedure, but still are providing a high degree of privacy to the user.

## Background

Tickets are used for an innumerable number of events: soccer matches, music festivals, exhibitions, etc. These tickets are ever more bought electronically. An increasing number of countries issue electronic identity cards to their citizens. Examples are Belgium, Estonia and Austria. These eID cards usually allow the holder to authenticate and to digitally sign documents, but often, they are very privacy unfriendly. For example, authentication using the Belgian eID card will usually lead to the divulgement of important personal data such as your national registration number (NRN). Despite these privacy dangers, the use of the eID card is promoted by the governments. We can thus expect that in the near future, electronic ticketing systems will arise based on the eID card. A trivial solution is easy to devise. However, this solution is not acceptable because it further endangers the card holder's privacy as profiles can easily be compiled, linked to each other and to the identity of the card holder. An advantage of the use of eID cards is that it is straightforward to impose restrictions on the maximum number of tickets that can be bought by one user, hence, thwarting sales on black markets. Sometimes, special offers are available for buyers under or over a certain age or living in the region where the event is organized. Here too, eID cards can help in securely conveying (proving) that these conditions are satisfied for the buyer. However, the use of these cards will usually disclose more information than is required.

For big events with thousands of attendants, the police would be helped if tickets were not anonymous, but could be linked to the identity of the attendants, or at least to the identity of the buyers of these tickets. Especially, when rows or riots occur, it would make it easier to identify and prosecute the instigators. However, the use of tickets attributable to individuals poses severe privacy risks and brings us closer to a "Big Brother" state.

This paper proposes two solutions where the eID card is needed to obtain an anonymized permit, allowing a user to obtain tickets in a privacy friendly way. The role of the eID card is thus reduced to a bootstrapping role. A first solution is based on pseudonym certificates, i.e. X.509 certificates containing a user's nym instead of a real identity. A second solution is based on the more enhanced anonymous credential systems, which allow to anonymously disclose only a subset of the personal attributes (or properties thereof) embedded in the credential. Both solutions are validated and compared with the trivial solution and with each other.

### Contributions

Two privacy preserving ticketing systems were proposed; one based on pseudonym certificates and one on anonymous credentials. This paper is a case study of an application where the citizen's privacy on the one hand and liabilities and restrictions on the other hand are important. This paper shows that these can be reconciled in a ticketing application using different techniques. Still the privacy unfriendly eID card is used as bootstrap.

## 5.7 Privacy-Enhanced Biometrics for e-ID Cards

- Claudia Diaz, and Pim Tuyls. *Privacy-enhanced biometrics for e-ID cards.* COSIC technical report, 10 pages, 2007.

### Summary

As identity documents migrate from paper to electronic formats, many countries are considering the possibility of including biometric authentication factors. In spite of this emerging trend, there is little scientific literature that addresses its privacy and security implications. This paper presents a study of the privacy and security issues that arise from the inclusion of biometric templates in e-ID cards, proposes the use of privacy-enhanced biometrics, and analyzes their privacy and security properties.

### Background

In the last years, many countries are moving from paper towards electronic identity documents (e-passports and e-IDs) for their citizens. As the issuing procedure usually involves reasonably secure identity checks, government-issued e-IDs present an opportunity for electronic authentication. Multi-factor authentication consists of combining different sources in order to ensure that someone is who he/she claims to be. Typically, it involves the possession of a physical authentication token (the e-ID), the knowledge of a secret (e.g., PIN number), and some biometric factor (e.g., fingerprint).

Physical tokens can be stolen, and secrets extracted through eavesdropping or coercion attacks. However, biometrics are linked to the physical presence of the subject they belong to and can not be forgotten or lost. For this reason, including biometric authentication factors in electronic IDs is being considered in several countries, such as France, The Netherlands, Portugal, Spain or Sweden. It is thus surprising how little attention has been paid in the scientific literature to the systematic study of the privacy and security issues that arise from such large-scale deployment of biometrics.

### Contributions

This paper investigates privacy and security threats that arise by introducing the inclusion of biometric templates in e-ID cards. We describe privacy-enhanced biometrics and discuss their advantages in the context of electronic documents. In particular, we present a solution for the integration of biometric authentication with e-ID cards, while guaranteeing the privacy of the citizens as much as possible.

## 5.8 Improvements for the Belgian eID Technology

- Pieter Verhaeghe, Jorn Lapon, Bart De Decker, Vincent Naessens and Kristof Verslype. *Security and Privacy Improvements for the Belgian eID Technology.* Proceedings of The Ifip Tc 11 24th International Information Security Conference. Pafos, (Cyprus), May 18-20, 2009.

### Summary

The Belgian Electronic Identity Card enables Belgian citizens to prove their identity digitally and to sign electronic documents. At the end of 2009, every Belgian citizen older than 12 years will have such an eID card. In the future, usage of the eID card may be mandatory. However, irresponsible use of the card may cause harm to individuals. Currently, there exist some privacy and security problems related to the use of the eID card. This paper focuses on solutions to tackle these problems. A new authentication protocol is introduced to substantially reduce the risk of abusing the single sign-on authentication and privacy friendly identity files are proposed to improve the citizen's privacy.

### Background

Belgium has introduced an electronic identity card [DWP06, DWP04] in 2002 as one of the first countries in Europe. The Belgian government aims at completing the roll-out by the end of 2009. At that time, each citizen will be the owner of an eID card. The card enables individuals to prove their identity digitally and to sign electronic documents. The Belgian eID card opens up new opportunities for the government, their citizens, service providers and application developers.

It is clear that many application developers benefit from this evolution. Today, integrating eID technology for authentication purposes is a real hot topic in Belgium. However, the usage of the eID card involves a few security and privacy hazards. Still, most citizens are unaware of these pitfalls, which is disturbing, since the usage of the card is highly encouraged both by the government and the industry.

## Contributions

The main contribution of this paper is the ***auth*** protocol and the privacy friendly identity files. The ***auth*** protocol tackles single sign-on authentication and trojan horses at the client side. The PFID files only disclose the necessary personal information and reduce linkabilities.

As a first step, the paper proposes to implement the ***auth*** protocol in software. The next step would be a new eID card that implements PFID-files and the ***auth*** protocol itself.

Although the storage efficiency of PFID-files is good, the current eID card has not enough free storage available to accomodate several domains. Hence, a smart card with more persistent storage will be necessary.

The privacy of the card holder can even further be enhanced by using anonymous credentials.

# Chapter 6

# Legal Aspects

## 6.1 Introduction

Technological developments have not stood still since the first roll-out of the Belgian eID card. Such developments also have a bearing on the legal obligations and liability of service providers. The data controller's security obligation under data protection regulations are measured against several factors. These factors include: cost, risk presented by the processing, nature of the information to be protected, and the state of the art. During this year's work plan we sought to analyze the specific privacy and security risks presented by the Belgian eID card in its current format and how they should be evaluated against the controller's security obligation.

Another relevant development is the increase in user-centric identity management applications. For instance, in the e-health domain we have seen the emergence of so-called 'personal' electronic health records, such as Google Health and Microsoft Health Vault. These applications claim to provide the user with a great deal of control and decision-making power as to which entities may be granted access to the information stored in his personal health record. Seeing as the user thus appears to obtain significant room to determine himself what should be the 'purposes and means' of the processing of his own personal data, it should be analyzed to what extent this affects the existing paradigms with regards to controllership and liability.

Finally, recent legislation in the telecommunications sector has indicated that anonymous communication services may be at odds with current policies supporting law enforcement's use of traffic and location data. It is therefore necessary to investigate whether the use or offering of such services is still permissible under the current framework, and to evaluate how

this affects the conception of a possible 'right to anonymity'.

### Relevance to ADAPID

New technical realities sometimes cause us to question how they will fit in the existing legal framework and whether the latter is well adapted to accommodate them. Conversely, the law also assigns consequences to certain design choices, and regulates the conditions within which technology may be deployed.

In [AC08]an analysis was made with regards the specific privacy and security risks presented by the Belgian eID card and how they must be seen in light of the controller's security obligation.

Ongoing research seeks to clarify whether existing data protection regulations and the roles defined therein are adequate to accommodate the new realities of data processing in advanced web applications (commonly referred to as "web 2.0"); whereby the user may play a substantial role in defining the ultimate purpose of a particular processing operation. The investigation of user-centric identity management applications is relevant to adapID as anonymous credentials are typically employed in user-centric applications.

The substantial increase in services enabling anonymous communication between two parties gives rise to the question as to whether such services are not in conflict with the European directive on data retention and their national implementations. Data retention requirements may also have an impact on how anonymity and data minimization will be construed in the future.

## 6.2 Due Processing of personal data in e-Government?

- Brendan Van Alsenoy and Danny De Cock, *Due processing of personal data in eGovernment? A Case Study of the Belgian electronic identity card.* Datenschutz und Datensicherheit, March 2008, 178-183.

### Summary

In this article, the authors evaluate the current authentication mechanisms for eGovernment employed in Belgium. It starts by providing a detailed overview of the eID card's specifications and functionalities, and then places it against the other means of authentication which are currently used.

After this analysis, a detailed account is given of the privacy risks related to the Belgian eID card. In first instance there is the issue of unrestricted access to the identity file. The identity file, address file, digital photo and citizen certificates are freely readable by any application that reads the eID

card chip. These files include personal and identifiable information, e.g., the card holder's national number, gender, her address, etc. This implies that a digital copy of this sensitive information may be more easily exposed whenever an eID card is used in an online or offline transaction.

The second component of the risk analysis relates to the issue of systematic and extended use of the same identifier across different identity management applications. To a greater or lesser extent, continuous reliance on the same identifier across contexts and sectors augments the following privacy threats:
- unlawful data exchange;
- loss of confidentiality;
- loss of "transactional" privacy;
- unlawful data aggregation;
- unlawful profiling and Knowledge Discovery in Databases (KDD);
- identity theft.

For some of these threats the link between identifier usage and the risk is quite strong, for others it is more removed. For instance, transactional privacy is probably the most directly threatened by recourse to the same identifier for every communication. Similarly, the risk of unlawful data aggregation increases greatly when the same identifier is continuously employed for attribute storage and communication. On the other hand, confidentiality is threatened more indirectly by systematic recourse to the same identifier. The most removed privacy risk is likely to be identity theft.

These risks are then placed against the controller's security obligation inscribed in art. 16 and 17 of Directive 95/46/EC [eu-95], which includes inter alia the obligation for the controller to take all reasonable measures "to prevent all other unlawful forms of processing" (art. 17). When determining whether this obligation has been appropriately met, two additional factors need to be taken into account. First, one must look at any other additional measures which may be in place to mitigate certain privacy risks (such as e.g. prior authorization requirements). One may also not lose sight of the fact that the security obligation of art. 17 is an 'obligation of means'. Cost and state-of-the-art are equally important elements to take into account when assessing whether the controller has adequately quitted himself of his obligations. At the time of the initial eID release, more privacy-friendly alternatives were in fact discussed, however, the maturity of these alternatives was considered insufficiently proven, and the infrastructure that would have been necessary to implement those alternatives had not yet been prepared. Today one should also take into account that an ever increasing amount of IdM architectures and applications have emerged in which uniform cross-sectoral identification is clearly no longer an interoperability requirement. The question is whether now, as the state-of-the art with respect to online transactions and interconnection between service providers has considerably evolved, a different approach will be sought.

### Contributions

The contribution of this paper is that it gives a more detailed account of the privacy risks associated with the Belgian eID card. In addition, it places these risks against the controller's security obligation under art. 17 of the Directive. This helps to situate the current version of the eID card in its historical context and to allow for new evaluation as the state of the art progresses.

## 6.3 Works in Progress

### 6.3.1 Is or will anonymity be a crime?

#### Summary

Enormous advances have been made in the research and development of privacy-enhancing technologies. Particularly in the field of electronic communications, there has been a huge boom in services which allow rendering communications anonymous or pseudonymous.

The European legislators have also concerned themselves with privacy in electronic communications. In 2002, the Directive concerning the processing of personal data and the protection of privacy in the electronic communications sector [eu-02] (also known as the "e-Privacy directive") was adopted. This directive imposes significant restrictions with regards to the further processing of traffic and location data. Four years later, the e-Privacy directive was amended by the Directive of 15 March 2006 'on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks' [eu-06b]. The so-called 'Data Retention Directive' mandates that Member States ensure the retention of data necessary to identify a.o. the source of a communication, its destination and the users' communication equipment (art. 5). These data are to be retained and held exclusively for law enforcement purposes.

The Belgian Law of 13 June 2005 on Electronic Communications [Tel05] already incorporated several provisions on data retention, particularly in art. 126-127. Art. 127, §2 states that the provision or use of a service or equipment which impedes the tracing and localization of private communications (or renders more these activities more difficult) is forbidden. Severe financial penalties are imposed in case of non-compliance. The law only provides exceptions for encryption systems that are used to guarantee the confidentiality of a communication or the security of electronic payments.

At first reading, it appears as if art. 127, §2 of the Law on Electronic Communications outlaws the offering and use of anonymous and pseudonymous communication technologies, seeing as the purpose of these tools is

exactly to render observation, tracing and localization more difficult or impossible. Further research must clarify the exact scope and purpose of this provision, and which types of anonymous communication methods are covered precisely. In addition, there is still some uncertainty as to whether or not this provision requires further implementation by Royal Decree in order for it to take effect. Finally, art. 127, §2 also appears to be at odds with several recommendations issued by the Council of Europe regarding the protection of personal data in the area of telecommunications services. [oE95]

### Contributions

This research will help to clarify the current regulatory framework governing the use of anonymous and pseudonymous communication services on both the national and EU level. It also seeks to determine whether there are less intrusive or otherwise more appropriate means of balancing the needs of law enforcement with the privacy interests of the individual.

## 6.3.2 Liability of the user in user-centric identity management

### Summary

Under the framework of Directive 95/46/EC, there are at least two actors implicated for every processing of personal data: a controller and a data subject. A data subject is any individual to whom the information relates, provided that he or she is identified or sufficiently identifiable (art. 2, a). The controller is the entity who alone, or jointly with others, determines the purposes and means of the processing. It is also possible that the controller chooses not to perform all the desired processing operations entirely by himself, but to have a whole or a part of the processing operations carried out by a different entity. A 'processor' is then an entity who carries out such operations on behalf of the data controller (art. 2, e).

At the time the Directive was adopted, the distinction between parties who control the processing of personal data and those who only process the data on behalf of another entity was relatively clear. [Kun07]The Directive assigns practically all responsibility for compliance with data protection regulations to the controller. Given the fundamental importance of the qualification as either a controller or a processor, it is crucial to be able to determine in which capacity an entity is performing a particular processing operation. Despite this reality, technological developments since the enactment of the Directive have made it increasingly difficult to apply the distinction between 'data controller' and 'data processor' in practice. [Kun07]

The emergence of web 2.0 and user-centric identity management applications have made it even more difficult to determine in which capacity an

entity is performing a specific processing operation. This is particularly the case for so-called 'personal' electronic health records, such as Google Health and Microsoft Health Vault. Many such applications claim to provide the user with a great deal of control and decision-making power as to which entities may access the stored information. The user thus appears to be given significant decision-making power as to both the 'purpose and means' of his personal health record.

The goal of this research is to investigate what roles and responsibilities the user assumes when employing such applications. A first matter which needs to be resolved is the question of whether an individual can be acting both as a controller and as a data subject. If so, it becomes important to know whether or not a service provider may restrict his role to 'processor', arguing that the application leaves the user with sufficient room in determining the purposes and means of his personal health record.

## Contributions

The outcome of this research should allow us to assess whether the current data protection regulations are well suited to govern user-centric identity management applications such as personal health records.

# Chapter 7

# Conclusions

In this deliverable we have summarized the main basic research results obtained in the second half of the project. We report on twenty pieces of work, most of which have been published in scientific journals or conferences, while others are still subject to submission or review processes. The main topics covered in this second research phase include privacy models and metrics, evaluation of anonymous communication systems, digital credentials, techniques for secure storage, secure and privacy enhanced applications based on the e-ID, and legal aspects of the use of the e-ID card.

In Chapter 2, we have shown how our work on digital credentials has a direct positive impact on the secure and privacy-preserving deployment of next-generation e-ID cards. We have first given an overview of the state of the art on digital credentials. Next, we have provided a list of the main privacy and security requirements desirable in a credential system, together with a number of possible application domains. We have then presented a number of novel research results that have been published by the partners. We are currently considering various extensions and improvements to previous results, in particular in the e-health domain.

The use of anonymous communications is key to achieve infrastructural privacy protection for any application that involves communication over the Internet. If no such precautions are taken, the protection provided by application-layer privacy enhancing technologies can be undermined by an attacker with access to the communications in the network. Furthermore, the lack of usable privacy models and metrics difficults the evaluation and comparison of existing technologies. In this document, we presented improvements in anonymity metrics such that it is possible to evaluate the loss of anonymity towards an attacker with access to extra information about the users of the system. We also described an attack on low-latency anonymity systems that improves significantly the state-of-the-art when deanonymizing messages sent through the network. This attack is particularly successful because it uses all information about the participants available to

the attacker, instead of considering them individually as suggested in previous work. Finally, we studied how combinatorial approaches can be used to evaluate anonymity networks, and showed how users that send and/or receive multiple messages reduce the anonymity provided by the system.

We have proposed several constructions for secure data storage in Chapter 4. Our solutions can be used to provide security and privacy in the deployment of databases and web servers intended to implement privacy-sensitive applications such as e-government, e-health or e-commerce. Further work should focus on strengthening security properties and on improving efficiency, as well as on providing frameworks and tools to integrate these privacy-preserving solutions into widely utilized technologies in the deployment of data storages, which will help to spread their implementation and use. Furthermore, in some of the schemes it is possible to improve usability aspects and to add extra features. For example, it would be interesting to permit more types of data searches in the scheme presented in Section 4.5, or to allow for more customer management features in the construction of Section 4.6.

In Chapter 5 we have presented several applications based on the Belgian e-ID that satisfy privacy and security requirements. The applications we propose fulfill similar functional requirements as privacy-invasive alternatives, while providing extra security and privacy protections. Some of the basic technologies studied in the project (e.g., digital credentials) have been key in enabling these applications. The diversity of the proposed applications (from e-petitions to ticketing systems to pay-as-you-drive insurance) illustrates the potential of authentication tokens such as the e-ID if combined with the appropriate security and privacy enhancing technologies.

The legal research has brought us to the conclusion that the Belgian e-ID presents specific privacy risks, several of which were a consequence of the state-of-the art at the time of deployment. Additional issues remain however. The propagation of the national number regardless of the relying party seems to unduly increase the risk of unlawful processing if the card is intended to be used outside the governmental sphere. Similarly, the issue of unrestricted access to the identity file remains difficult to reconcile with the principle of data minimization. The research concerning anonymous communication services performed so far indicates that the use or offering of particular types of these services is likely to violate Belgian law, despite the concerns expressed by the Privacy Commission. Whether or not this situation is in line with policy at the European level remains to be determined. Finally, the current legal framework for data protection might be need to be revised to accommodate the new realities of web 2.0 and the role of users as possible data controllers. It has become increasingly difficult to qualify the roles of the different actors involved, which do not always correspond unambiguously with the traditional controller-processor distinction within the current framework.

# Bibliography


[AC08] Brendan Van Alsenoy and Danny De Cock. Due processing of personal data in egovernment? a case study of the belgian electronic identity card. *Datenschutz und Datensicherheit*, pages 178–183, March 2008.

[ACJT00] Giuseppe Ateniese, Jan Camenisch, Marc Joye, and Gene Tsudik. A practical and provably secure coalition-resistant group signature scheme. In Mihir Bellare, editor, *CRYPTO 2000*, volume 1880 of *LNCS*, pages 255–270, 2000.

[Aio] Aioi. `http://www.ioi-sonpo.co.jp/`.

[AIR01] William Aiello, Yuval Ishai, and Omer Reingold. Priced oblivious transfer: How to sell digital goods. In Birgit Pfitzmann, editor, *EUROCRYPT*, volume 2045 of *Lecture Notes in Computer Science*, pages 119–135. Springer, 2001.

[AK03a] D. Agrawal and D. Kesdogan. Measuring anonymity: The disclosure attack. *IEEE Security & Privacy*, 1(6):27–34, 2003.

[AK03b] Dakshi Agrawal and Dogan Kesdogan. Measuring anonymity: The disclosure attack. *IEEE Security & Privacy*, 1(6):27–34, 2003.

[ANS98] Ross J. Anderson, Roger M. Needham, and Adi Shamir. The steganographic file system. In *Proceedings of the Second International Workshop on Information Hiding*, LNCS 1525, pages 73–82. Springer-Verlag, 1998.

[Ass] American Automobile Association. `http://www.aaa.com/`.

[Ass98] National Motorist Association. NMA's position on auto insurance. `http://www.motorists.org`, 1998.

[BB04] Dan Boneh and Xavier Boyen. Efficient selective-id secure identity-based encryption without random oracles. In *EUROCRYPT*, pages 223–238, 2004.

[BBS04a] Dan Boneh, Xavier Boyen, and Hovav Shacham. Short group signatures. In Matthew K. Franklin, editor, *CRYPTO*, volume 3152 of *Lecture Notes in Computer Science*, pages 41–55. Springer, 2004.

[BBS04b] Dan Boneh, Xavier Boyen, and Hovav Shacham. Short group signatures using strong Diffie-Hellman. In *CRYPTO*, volume 3152 of LNCS, pages 41–55, 2004.

[BCC04] Ernie Brickell, Jan Camenisch, and Liqun Chen. Direct anonymous attestation. Technical Report Research Report RZ 3450, IBM Research Division, March 2004.

[BCKL08a] Mira Belenkiy, Melissa Chase, Markulf Kohlweiss, and Anna Lysyanskaya. Compact e-cash and simulatable vrfs revisited, 2008.

[BCKL08b] Mira Belenkiy, Melissa Chase, Markulf Kohlweiss, and Anna Lysyanskaya. P-signatures and noninteractive anonymous credentials. In Ran Canetti, editor, *TCC*, volume 4948 of *Lecture Notes in Computer Science*, pages 356–374. Springer, 2008.

[BDCOP04] D. Boneh, G. Di Crescenzo, R. Ostrovsky, and G. Persiano. Public key encryption with keyword search. *Proceedings of Eurocrypt*, 4, 2004.

[BDJB04] A. Jerman Blazic and B. Dzonova-Jerman-Blazic. Implementing trustworthy internet based long term electronic preservation service – the eKeeper project. In M. H. Hamza, editor, *IASTED International Conference on Communications, Internet, and Information Technology*, pages 291–296, 2004.

[BDMP91] Manuel Blum, Alfredo De Santis, Silvio Micali, and Guiseppe Persiano. Non-interactive zero-knowledge. *SIAM Journal of Computing*, 20(6):1084–1118, 1991.

[BF01] Dan Boneh and Matthew K. Franklin. Identity-based encryption from the weil pairing. In Joe Kilian, editor, *CRYPTO*, volume 2139 of *Lecture Notes in Computer Science*, pages 213–229. Springer, 2001.

[BFK00] Oliver Berthold, Hannes Federrath, and Marit Köhntopp. Project "anonymity and unobservability in the internet". In *CFP '00: Proceedings of the tenth conference on Computers, freedom and privacy*, pages 57–65, New York, NY, USA, 2000. ACM.

[BFK01] Oliver Berthold, Hannes Federrath, and Stefan Köpsell. Web mixes: A system for anonymous and unobservable internet access. In *Designing Privacy Enhancing Technologies*, pages 115–129. Springer-Verlag, LNCS 2009, 2001.

[BFM88] Manuel Blum, Paul Feldman, and Silvio Micali. Non-interactive zero-knowledge and its applications (extended abstract). In *Proceedings of the Twentieth Annual ACM Symposium on Theory of Computing*, pages 103–112, Chicago, Illinois, 2–4 May 1988.

[Bla07] A. Jerman Blazic. Long term trusted archive services. In *First International Conference on the Digital Society ICDS*, page 29. IEEE Computer Society, Jan 2007.

[BLS01] Dan Boneh, Ben Lynn, and Hovav Shacham. Short signatures from the weil pairing. In Colin Boyd, editor, *ASIACRYPT*, volume 2248 of *Lecture Notes in Computer Science*, pages 514–532. Springer, 2001.

[Bra93] Stefan Brands. An efficient off-line electronic cash system based on the representation problem. Technical Report CS-R9323, CWI, April 1993.

[Bra94] Stefan Brands. Untraceable off-line cash in wallet with observers. In *Advances in cryptology – Crypto'93*, volume 773 of *LNCS*, pages 302–318. Springer-Verlag, 1994.

[Bra99] Stefan Brands. *Rethinking Public Key Infrastructure and Digital Certificates— Building in Privacy.* PhD thesis, Eindhoven Inst. of Tech. The Netherlands, 1999.

[Bra00] Stefan Brands. *Rethinking Public Key Infrastructures and Digital Certificates: Building in Privacy.* The MIT Press, 2000.

[BS05] A. Jerman Blazic and P. Sylvester. Provision of long-term archiving service for digitally signed documents using an archive interaction protocol. In D. W. Chadwick and G. Zhao, editors, *EuroPKI*, pages 240–254, 2005.

[BW06a] Xavier Boyen and Brent Waters. Anonymous hierarchical identity-based encryption (without random oracles). In *CRYPTO*, pages 290–307, 2006.

[BW06b] Xavier Boyen and Brent Waters. Compact group signatures without random oracles. In *EUROCRYPT*, pages 427–444, 2006.

[BW07] Xavier Boyen and Brent Waters. Full-domain subgroup hiding and constant-size group signatures. In *Public Key Cryptography*, pages 1–15, 2007.

[Can01] R. Canetti. Universally composable security: A new paradigm for cryptographic protocols. In *FOCS '01: Proceedings of the 42nd IEEE symposium on Foundations of Computer Science*, page 136, Washington, DC, USA, 2001. IEEE Computer Society.

[CFN90] David Chaum, Amos Fiat, and Moni Naor. Untraceable electronic cash. In *CRYPTO '90*, volume 403 of *LNCS*, pages 319–327, 1990.

[CGH04a] Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *J. ACM*, 51(4):557–594, 2004.

[CGH04b] Ran Canetti, Oded Goldreich, and Shai Halevi. The random oracle methodology, revisited. *J. ACM*, 51(4):557–594, 2004.

[CGH06] Sébastien Canard, Aline Gouget, and Emeline Hufschmitt. A handy multi-coupon system. In Jianying Zhou, Moti Yung, and Feng Bao, editors, *ACNS*, volume 3989 of *Lecture Notes in Computer Science*, pages 66–81, 2006.

[CGKS95] Benny Chor, Oded Goldreich, Eyal Kushilevitz, and Madhu Sudan. Private information retrieval. In *FOCS*, pages 41–50, 1995.

[CGN98] Benny Chor, Niv Gilboa, and Moni Naor. Private information retrieval by keywords, 1998.

[Cha81a] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, February 1981.

[Cha81b] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 4(2):84–88, 1981.

[Cha82] David Chaum. Blind signatures for untraceable payments. In *CRYPTO '82*, pages 199–203. Plenum Press, 1982.

[Cha85a] David Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, October 1985.

[Cha85b] David Chaum. Security without identification: transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, 1985.

[Cha07] Rafik Chaabouni. Efficient protocols for set membership and range proofs. Technical report, EPFL LASEC, IBM ZRL, October 2007.

[CHK+06a] Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya, and Mira Meyerovich. How to win the clonewars: efficient periodic n-times anonymous authentication. In Juels et al. [JWdV06], pages 201–210.

[CHK+06b] Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya, and Mira Meyerovich. How to win the clonewars: efficient periodic n-times anonymous authentication. In *CCS '06: Proceedings of the 13th ACM conference on Computer and communications security*, pages 201–210, New York, NY, USA, 2006. ACM Press.

[CHK+06c] Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya, and Mira Meyerovich. How to win the clonewars: efficient periodic n-times anonymous authentication. In Juels et al. [JWdV06], pages 201–210.

[CHL05] Jan Camenisch, Susan Hohenberger, and Anna Lysyanskaya. Compact E-Cash. In *EUROCRYPT*, volume 3494 of LNCS, pages 302–321, 2005.

[CHL06] Jan Camenisch, Susan Hohenberger, and Anna Lysyanskaya. Balancing accountability and privacy using e-cash. In *SCN (to appear)*, 2006.

[CL01a] Jan Camenisch and Anna Lysyanskaya. Efficient non-transferable anonymous multi-show credential system with optional anonymity revocation. In *Advances in Cryptology – EuroCrypt'01*, volume 2045 of *LNCS*, pages 93–118. Springer Verlag, 2001.

[CL01b] Jan Camenisch and Anna Lysyanskaya. Efficient non-transferable anonymous multi-show credential system with optional anonymity revocation. In Birgit Pfitzmann, editor, *EUROCRYPT 2001*, volume 2045 of *LNCS*, pages 93–118. Springer Verlag, 2001.

[CL01c] Jan Camenisch and Anna Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional ano-

nymity revocation. In *Advances in Cryptology - EUROCRYPT 2001*, volume LNCS 2045, pages 93–118. Springer, 2001.

[CL02] Jan Camenisch and Anna Lysyanskaya. A signature scheme with efficient protocols. In *SCN 2002*, volume 2576 of *LNCS*, pages 268–289, 2002.

[CL04a] Jan Camenisch and Anna Lysyanskaya. Signature schemes and anonymous credentials from bilinear maps. In *Advances in cryptology – Crypto'04*, volume 3152 of *LNCS*, pages 56–72. Springer-Verlag, 2004.

[CL04b] Jan Camenisch and Anna Lysyanskaya. Signature schemes and anonymous credentials from bilinear maps. In *CRYPTO 2004*, volume 3152 of *LNCS*, pages 56–72, 2004.

[Cla06] Sebastian Clauß. A framework for quantification of linkability within a privacy-enhancing identity management system. In *Emerging Trends in Information and Communication Security*, pages 191–205. Springer, LNCS 3995, 2006.

[CLM07] Jan Camenisch, Anna Lysyanskaya, and Mira Meyerovich. Endorsed e-cash. In *IEEE Symposium on Security and Privacy*, pages 101–115. IEEE Computer Society, 2007.

[CNS07] Jan Camenisch, Gregory Neven, and Abhi Shelat. Simulatable adaptive oblivious transfer. In Moni Naor, editor, *EUROCRYPT*, volume 4515 of *Lecture Notes in Computer Science*, pages 573–590. Springer, 2007.

[Coc] Danny De Cock. Non-official information on the Belgian Electronic Personal Identification Card. `https://www.cosic.esat.kuleuven.be/belpic/`.

[Cor] Toyota Motor Corporation. `http://www.toyota.co.jp/`.

[CP92] David Chaum and Torben P. Pedersen. Wallet databases with observers. In *Advances in cryptology – Crypto'92*, volume 740 of *LNCS*, pages 89–105. Springer-Verlag, 1992.

[CP93] David Chaum and Torben Pryds Pedersen. Transferred cash grows in size. In *EUROCRYPT '92*, volume 658 of *LNCS*, pages 390–407, 1993.

[CS97] Jan Camenisch and Markus Stadler. Efficient group signature schemes for large groups. In Burt Kaliski, editor, *CRYPTO '97*, volume 1296 of *LNCS*, pages 410–424. Springer Verlag, 1997.

[CS05] Sanjit Chatterjee and Palash Sarkar. Trading time for space: Towards an efficient ibe scheme with short(er) public parameters in the standard model. In *ICISC*, pages 424–440, 2005.

[CS06] Sebastian Clauß and Stefan Schiffner. Structuring anonymity metrics. *Proceedings of the ACM Workshop on Digital Identity Management*, pages 55–62, 2006.

[CvH91] David Chaum and Eugène van Heyst. Group signatures. In Donald W. Davies, editor, *EUROCRYPT '91*, volume 547 of *LNCS*, pages 257–265. Springer-Verlag, 1991.

[CVH02] Jan Camenisch and Els Van Herreweghen. Design and implementation of the *idemix* anonymous credential system. In *Proc. 9th ACM Conference on Computer and Communications Security*. acm press, 2002.

[CW04] S. Chokhani and C. Wallace. Trusted archiving. In *Proceedings of the 3rd Annual PKI R&D Workshop*. NIST, April 2004.

[Dam90] Ivan Bjerre Damgård. Payment systems and credential mechanism with provable security against abuse by individuals. In Shafi Goldwasser, editor, *CRYPTO '88*, volume 403 of *LNCS*, pages 328–335. Springer Verlag, 1990.

[Dan03] G. Danezis. Statistical disclosure attacks: Traffic confirmation in open environments. In Gritzalis, Vimercati, Samarati, and Katsikas, editors, *Proceedings of Security and Privacy in the Age of Uncertainty, (SEC2003)*, pages 421–426, Athens, May 2003. IFIP TC11, Kluwer.

[Dan04] George Danezis. The traffic analysis of continuous-time mixes. In *Proceedings of Privacy Enhancing Technologies workshop (PET 2004)*, volume 3424 of *LNCS*, pages 35–50, May 2004.

[DCSP02] Claudia Diaz, Joris Claessens, Stefaan Seys, and Bart Preneel. Information theory and anonymity. In B. Macq and J.-J. Quisquater, editors, *Werkgemeenschap voor Informatie en Communicatietheorie*, pages 179–186, 2002.

[DDM03a] G. Danezis, R. Dingledine, and N. Mathewson. Mixminion: Design of a Type III Anonymous Remailer Protocol. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, pages 2–15, May 2003.

[DDM03b] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a type iii anonymous remailer proto-

col. In *Proceedings of the 2003 IEEE Symposium on Security and Privacy*, pages 2–15, 2003.

[DDP06] Ivan Damgård, Kasper Dupont, and Michael Østergaard Pedersen. Unclonable group identification. In *EUROCRYPT '06*, pages 555–572, 2006.

[DDT07] G. Danezis, C. Diaz, and C. Troncoso. Two-sided statistical disclosure attack. In Nikita Borisov and Philippe Golle, editors, *Proceedings of Privacy Enhancing Technologies, 7th International Workshop, PET 2007*, volume 4776 of *Lecture Notes in Computer Science*, page 15, Ottawa,Canada, 2007. Springer-Verlag.

[DMS04] Roger Dingledine, Nick Mathewson, and Paul Syverson. Tor: The second-generation onion router. In *Proceedings of the 13th USENIX Security Symposium*, pages 303–320. USENIX, 2004.

[DNO08] Ivan Damgrd, Jesper Buus Nielsen, and Claudio Orlandi. Essentially optimal universally composable oblivious transfer. Cryptology ePrint Archive, Report 2008/220, 2008. `http://eprint.iacr.org/`.

[DNRS03] Cynthia Dwork, Moni Naor, Omer Reingold, and Larry J. Stockmeyer. Magic functions. *J. ACM*, 50(6):852–921, 2003.

[DP04] Claudia Diaz and Bart Preneel. Reasoning about the anonymity provided by pool mixes that generate dummy traffic. In *Information Hiding*, pages 309–325. Springer, LNCS 3200, 2004.

[DPW06] Yuxin Deng, Jun Pang, and Peng Wu. Measuring anonymity with relative entropy. In Theodosis Dimitrakos, Fabio Martinelli, Peter Y. A. Ryan, and Steve A. Schneider, editors, *Formal Aspects in Security and Trust*, pages 65–79. Springer, LNCS 4691, 2006.

[DS03] Claudia Diaz and Andrei Serjantov. Generalising mixes. In *Designing Privacy Enhancing Technologies, Proceedings of PET'03*, pages 18–31. Springer-Verlag, LNCS 2760, 2003.

[DSCP02] Claudia Diaz, Stefaan Seys, Joris Claessens, and Bart Preneel. Towards measuring anonymity. In *Designing Privacy Enhancing Technologies, Proceedings of PET'02*, pages 54–68. Springer-Verlag, LNCS 2482, 2002.

[DSMP88] Alfredo De Santis, Silvio Micali, and Giuseppe Persiano. Non-interactive zero-knowledge proof systems. In Carl Pomerance, editor, *CRYPTO '87*, volume 293 of *LNCS*, pages 52–72. Springer-Verlag, 1988.

[DTS08] Claudia Diaz, Carmela Troncoso, and Andrei Serjantov. On the impact of social network profiling on anonymity. In Nikita Borisov and Ian Goldberg, editors, *Privacy Enhancing Technologies*, page 19. Springer, LNCS (in print), 2008.

[DWP04] Danny De Cock, Karel Wouters, and Bart Preneel. Introduction to the Belgian EID Card: BELPIC. In Stefanos Gritzalis, Sokratis K. Katsikas, and J. Lopez, editors, *European PKI Workshop: Research and Applications*, volume 3093 of *Lecture Notes in Computer Science*, pages 1–13, Samos Island,GR, 2004. Springer-Verlag.

[DWP06] Danny De Cock, Christopher Wolf, and Bart Preneel. The Belgian Electronic Identity Card (Overview). In Jana Dittmann, editor, *Sicherheit 2005: Sicherheit - Schutz und Zuverlässigkeit, Beiträge der 3rd Jahrestagung des Fachbereichs Sicherheit der Gesellschaft für Informatik e.v. (GI)*, volume LNI P-77 of *Lecture Notes in Informatics (LNI)*, pages 298–301, Magdeburg,DE, 2006. Bonner Köllen Verlag.

[EAFH04] Fernando Esponda, Elena S. Ackley, Stephanie Forrest, and Paul Helman. Online negative databases. In *ICARIS*, pages 175–188, 2004.

[EAH+06] Fernando Esponda, Elena S. Ackley, Paul Helman, Haixia Jia, and Stephanie Forrest. Protecting data privacy through hard-to-reverse negative databases. In *ISC*, pages 72–84, 2006.

[eco06] The non-denial of the non-self. *The Economist*, August 2006.

[EJA+04] Endre Bangerter, Jan Camenisch, Anna Lysyanskaya. In, Cambridge 12th International Workshop on Security Protocols 2004, and 2004. England, 26-28 April 2004. Springer Verlag. A Cryptographic Framework for the Controlled Release Of Certified Data. In *12th International Workshop on Security Protocols 2004*, Cambridge, England, 26 April 2004. Springer.

[ESY07] Matthew Edman, Fikret Sivrikaya, and Bulent Yener. A combinatorial approach to measuring anonymity. *Intelligence and Security Informatics, 2007 IEEE*, pages 356–363, 2007.

[eu-95] Directive 95/46/ec of the european parliament and of the council of 24 october 1995 on the protection of individuals with regard to the processing of personal data and on the free movement of such data (eu data protection directive). Official Journal of the European Union, November 1995.

[eu-02] Directive 2002/58/ec of the european parliament and of the council of 12 july 2002 on the protection of privacy in the electronic communications sector (directive on privacy and electronic communications). Official Journal of the European Union, July 2002.

[eu-06a] Directive 2006/24/ec of the european parliament and of the council. Official Journal of the European Union, April 2006.

[eu-06b] Directive 2006/24/ec of the european parliament and of the council of 15 march 2006 on the retention of data generated or processed in connection with the provision of publicly available electronic communications services or of public communications networks and amending directive 2002/58/ec. Official Journal of the European Union, April 2006.

[FMP07] Matthias Franz, Bernd Meyer, and Andreas Pashalidis. Attacking unlinkability: The importance of context. In Nikita Borisov and Philippe Golle, editors, *Privacy Enhancing Technologies*, pages 1–16. Springer, LNCS 4776, 2007.

[FO98] E. Fujisaki and T. Okamoto. A practical and provably secure scheme for publicly verifiable secret sharing and its applications. In Kaisa Nyberg, editor, *Advances in Cryptology – EUROCRYPT '98*, volume 1403 of LNCS, pages 32–46. Springer, 1998.

[FS87] Amos Fiat and Adi Shamir. How to prove yourself: Practical solutions to identification and signature problems. In Andrew M. Odlyzko, editor, *CRYPTO '86*, volume 263 of *LNCS*, pages 186–194. Springer Verlag, 1987.

[FY93] Matthew Franklin and Moti Yung. Towards provably secure efficient electronic cash. In *proceedings of ICALP '93*, volume 700 of LNCS, pages 265–276, 1993.

[GA04] Rüdiger Grimm and Patrick Aichroth. Privacy protection for signed media files: a separation-of-duty approach to the lightweight drm (lwdrm) system. In Jana Dittmann and Jessica J. Fridrich, editors, *MM&Sec*, pages 93–99. ACM, 2004.

[Gen06] C. Gentry. Practical identity-based encryption without random oracles. *EUROCRYPT*, pages 445–464, 2006.

[GH07a] Matthew Green and Susan Hohenberger. Blind identity-based encryption and simulatable oblivious transfer. In *ASIACRYPT*, pages 265–282, 2007.

[GH07b] Matthew Green and Susan Hohenberger. Blind identity-based encryption and simulatable oblivious transfer. In *ASIACRYPT*, pages 265–282, 2007.

[GH08] Matthew Green and Susan Hohenberger. Universally composable adaptive oblivious transfer. Cryptology ePrint Archive, Report 2008/163, 2008. `http://eprint.iacr.org/`.

[GK03a] Shafi Goldwasser and Yael Tauman Kalai. On the (in)security of the Fiat-Shamir paradigm. In *Proc. 44th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 102–115. IEEE Computer Society Press, 2003.

[GK03b] Shafi Goldwasser and Yael Tauman Kalai. On the (in)security of the fiat-shamir paradigm. In *FOCS*, pages 102–, 2003.

[GL04] Charles Giefer and Julie Letchner. Mojitos: A distributed steganographic file system. Technical Report, University of Washington, 2004.

[GMW86] Oded Goldreich, Silvio Micali, and Avi Wigderson. Proofs that yield nothing but their validity and a method of cryptographic protocol design. In *Proc. 27th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 174–187. IEEE Computer Society Press, 1986.

[GRS96] David Goldschlag, Michael Reed, and Paul Syverson. Hiding routing information. In *Information Hiding*, pages 137–150, 1996.

[GS07] Jens Groth and Amit Sahai. Efficient non-interactive proof systems for bilinear groups. `http://eprint.iacr.org/2007/155`, 2007.

[GS08] Jens Groth and Amit Sahai. Efficient non-interactive proof systems for bilinear groups. In Nigel P. Smart, editor, *EUROCRYPT*, volume 4965 of *Lecture Notes in Computer Science*, pages 415–432. Springer, 2008.

[Her08] Javier Herranz. Restricted adaptive oblivious transfer. Cryptology ePrint Archive, Report 2008/182, 2008. `http://eprint.iacr.org/`.

[Hig] The Higgins Trust Framework Project. http://www.eclipse.org/higgins/. link functional as of Feb 2007.

[HK06] S. Haber and P. Kamat. A content integrity service for long-term digital archives. In *IS&T Archiving Conference (Archiving 2006)*, volume 3, pages 159–164. The Society for Imaging Science and Technology, May 2006.

[HR02] Steven Hand and Timothy Roscoe. Mnemosyne: Peer-to-peer steganographic storage. In *Revised Papers from the First International Workshop on Peer-to-Peer Systems*, LNCS 2429, pages 130–140. Springer-Verlag, 2002.

[Ide] The Identity Mixer. http://www.zurich.ibm.com/security/idemix/. link functional as of Feb 2007.

[IK97] Yuval Ishai and Eyal Kushilevitz. Private simultaneous messages protocols with applications. In *In Proc. of 5th ISTCS*, pages 174–183, 1997.

[Ins] Hollard Insurance. `http://www.payasyoudrive.co.za/`.

[IT05] ITU-T. Information technology - open systems interconnection - the directory: Public-key and attribute certificate frameworks – X.509 Recommendation, 2005. `http://www.itu.int/rec/T-REC-X.509/en`.

[JS04] Stanislaw Jarecki and Vitaly Shmatikov. Handcuffing big brother: an abuse-resilient transaction escrow scheme. In *EUROCRYPT*, volume 3027 of *LNCS*, pages 590–608, 2004.

[JWdV06] Ari Juels, Rebecca N. Wright, and Sabrina De Capitani di Vimercati, editors. *Proceedings of the 13th ACM Conference on Computer and Communications Security, CCS 2006, Alexandria, VA, USA, Ioctober 30 - November 3, 2006*. ACM, 2006.

[Kun07] C. Kuner, 2007.

[KW99] P. Koargonkar and L. Wolin. A multivariate analysis of web usage. *Journal of Advertising Research*, pages 53–68, March/April 1999.

[LGGG07] Michael LeMay, George Gross, Carl Gunter, and Sanjam Garg. Unified architecture for large-scale attested metering. In *Proceedings of the 40th Annual Hawaii International Conference on System Sciences*, 2007.

[Lip05] Helger Lipmaa. An oblivious transfer protocol with log-squared communication. In *Proceedings of the 8th International Information Security Conference*, volume 3650 of *LNCS*, pages 314–328. Springer-Verlag, 2005.

[Lit07] Todd Litman. Distance-based vehicle insurance feasibility, costs and benefits. Technical report, Victoria Transport Policy Institute, 2007.

[LOL04] Deok-Gyu Lee, Hyung-Geun Oh, and Im-Yeong Lee. A study on contents distribution using electronic cash system. In *EEE '04: Proceedings of the 2004 IEEE International Conference on e-Technology, e-Commerce and e-Service (EEE'04)*, pages 333–340, Washington, DC, USA, 2004. IEEE Computer Society.

[LRSW99] Anna Lysyanskaya, Ron Rivest, Amit Sahai, and Stefan Wolf. Pseudonym systems. In Howard Heys and Carlisle Adams, editors, *Selected Areas in Cryptography*, volume 1758 of *LNCS*, 1999.

[Lys02] Anna Lysyanskaya. *Signature Schemes and Applications to Cryptographic Protocol Design.* PhD thesis, Massachusetts Institute of Technology, Cambridge, Massachusetts, September 2002.

[McA07] J. McAdams. 27 billion gigabytes to be archived by 2010. *Computerworld*, December 2007.

[MCPS03a] U. Möller, L. Cottrell, P. Palfrader, and L. Sassaman. Mixmaster Protocol — Version 2. IETF Internet Draft, July 2003.

[MCPS03b] Ulf Moller, Lance Cottrel, Peter Palfrader, and Len Sassaman. Mixmaster protocol - version 2. `http://www.abditum.com/mixmaster-spec.txt`, 2003.

[MK99] Andrew D. McDonald and Markus G. Kuhn. Stegfs: A steganographic file system for linux. In *Proceedings of the Third International Workshop on Information Hiding*, LNCS 1768, pages 462–477. Springer-Verlag, 1999.

[Moo98] G. E. Moore. Cramming more components onto integrated circuits. *Proceedings of the IEEE*, 86(1):82–85, 1998.

[MSC08] Microsoft's credentica purchase helps it sprint ahead of openid. NetworkWorld, 19 Mar 2008. `http://www.networkworld.com/newsletters/dir/2008/0317id2.html`.

[mShWfH] Hung min Sun, King hang Wang, and Chi fu Hung. Towards privacy preserving digital rights management using oblivious transfer.

[Nac07] D. Naccache. Secure and practical identity-based encryption. *Information Security, IET*, 1(2):59–64, 2007.

[NP99] Moni Naor and Benny Pinkas. Oblivious transfer with adaptive queries. In Michael J. Wiener, editor, *CRYPTO*, volume 1666 of *Lecture Notes in Computer Science*, pages 573–590. Springer, 1999.

[NSN05] Lan Nguyen and Rei Safavi-Naini. Dynamic $k$-times anonymous authentication. In *ACNS 2005*, number 3531 in LNCS, pages 318–333. Springer Verlag, 2005.

[oE95] Council of Europe. Recommendation no. r(95)4 of the committee of ministers to member states on the protection of personal data in the area of telecommunication services, with particular reference to telephone services. Adopted by the Committe of Ministers on 7 February 1995 at the 528th Meeting of the Minister's Deputies, February 1995.

[OI07] Rafail Ostrovsky and William E. Skeith III. Private searching on streaming data. *J. Cryptology*, 20(4):397–430, 2007.

[OK04] Wakaha Ogata and Kaoru Kurosawa. Oblivious keyword search. *J. Complexity*, 20(2-3):356–371, 2004.

[OU98] Tatsuaki Okamoto and Shigenori Uchiyama. A new public-key cryptosystem as secure as factoring. In *Advances in Cryptology – EUROCRYPT'98*, volume 1403 of *LNCS*, pages 308–318. Springer-Verlag, 1998.

[Ped92] Torben Pryds Pedersen. Non-interactive and information-theoretic secure verifiable secret sharing. In *CRYPTO '92*, volume 576 of *LNCS*, pages 129–140, 1992.

[PH01] Andreas Pfitzmann and Marit Hansen. Anonymity, unobservability and pseudonymity – a proposal for terminology. In *Designing Privacy Enhancing Technologies, Proceedings of PET'00*, pages 1–9. Springer-Verlag, LNCS 2009, 2001.

[Ŕ61] Alfred Rényi. On measures of entropy and information. *Proceedings of the Fourth Berkeley Symposium Mathematical Statistics and Probability*, 1:547–561, 1961.

[Rab81] Michael O. Rabin. How to exchange secrets by oblivious transfer. 1981.

[RSG98] Michael Reed, Paul Syverson, and David Goldschlag. Anonymous Connections and Onion Routing. *IEEE Journal on Selected Areas in Communications*, 16(4):482–494, 1998.

[SD02a] A. Serjantov and G. Danezis. Towards an information theoretic metric for anonymity. In Roger Dingledine and Paul Syverson, editors, *Proceedings of Privacy Enhancing Technologies Workshop (PET 2002)*. Springer-Verlag, LNCS 2482, April 2002.

[SD02b] Andrei Serjantov and George Danezis. Towards an information theoretic metric for anonymity. In *Designing Privacy Enhancing Technologies, Proceedings of PET'02*, pages 41–53. Springer-Verlag, LNCS 2482, 2002.

[SDS02] Andrei Serjantov, Roger Dingledine, and Paul Syverson. From a trickle to a flood: Active attacks on several mix types. In Fabien Petitcolas, editor, *Proceedings of Information Hiding Workshop (IH 2002)*. Springer-Verlag, LNCS 2578, October 2002.

[Sha48] Claude Shannon. A mathematical theory of communication. *The Bell System Technical Journal*, 27:379–423:623–656, 1948.

[SSR08] Bhavani Shankar, Kannan Srinathan, and C. Pandu Rangan. Alternative protocols for generalized oblivious transfer. In Shrisha Rao, Mainak Chatterjee, Prasad Jayanti, C. Siva Ram Murthy, and Sanjoy Kumar Saha, editors, *ICDCN*, volume 4904 of *Lecture Notes in Computer Science*, pages 304–309. Springer, 2008.

[TDDP07] Carmela Troncoso, Claudia Diaz, Orr Dunkelman, and Bart Preneel. Traffic analysis attacks on a continuously-observable steganographic file system. In T. Furon et al., editor, *Information Hiding, 9th International Workshop*, volume 4567 of *Lecture Notes in Computer Science*, pages 220–236, Saint-Malo,FR, 2007. Springer-Verlag.

[TECA07] Janice Tsai, Serge Egelman, Lorrie Cranor, and Ro Acquisti. The effect of online privacy information on purchasing behavior: An experimental study, working paper. june 2007.

[Tel05] Law of 13 juni 2005 concerning electronic communications. Belgian State Gazette, June 2005.

[TFS04] Isamu Teranishi, Jun Furukawa, and Kazue Sako. k-times anonymous authentication (extended abstract). In Pil Joong Lee, editor, *ASIACRYPT*, volume 3329 of *Lecture Notes in Computer Science*, pages 308–322. Springer, 2004.

[THV04] Gergely Tóth, Zoltán Hornák, and Ferenc Vajda. Measuring anonymity revisited. In *Proceedings of the Ninth Nordic Workshop on Secure IT Systems*, pages 85–90, 2004.

[Tob03] Christian Tobias. Practical oblivious transfer protocols. In *IH '02: Revised Papers from the 5th International Workshop on Information Hiding*, pages 415–426, London, UK, 2003. Springer-Verlag.

[TS06] Isamu Teranishi and Kazue Sako. -times anonymous authentication with a constant proving cost. In Moti Yung, Yevgeniy Dodis, Aggelos Kiayias, and Tal Malkin, editors, *Public Key Cryptography*, volume 3958 of *Lecture Notes in Computer Science*, pages 525–542. Springer, 2006.

[Uni] Norwich Union. `http://www.norwichunion.com/pay-as-you-drive/`.

[Wag08] David Wagner, editor. *Advances in Cryptology - CRYPTO 2008, 28th Annual International Cryptology Conference, Santa Barbara, CA, USA, August 17-21, 2008. Proceedings*, volume 5157 of *Lecture Notes in Computer Science*. Springer, 2008.

[WBDS04] Brent R. Waters, Dirk Balfanz, Glenn Durfee, and Diana K. Smetters. Building an encrypted and searchable audit log. In *NDSS*, 2004.

[Wei05] Victor K. Wei. More compact e-cash with efficient coin tracing. Cryptology ePrint Archive, Report 2005/411, 2005. `http://eprint.iacr.org/`.

[Yao86] Andrew C. Yao. How to generate and exchange secrets. In *Proc. 27th IEEE Symposium on Foundations of Computer Science (FOCS)*, pages 162–167, 1986.

[ZB04] Fayyaz Zahid and Craig Barton. Pay per mile insurance. Technical report, Davenport University, 2004.

[ZB05] Ye Zhu and Riccardo Bettati. Anonymity vs. information leakage in anonymity systems. In *ICDCS '05: Proceedings of the 25th IEEE International Conference on Distributed Computing Systems (ICDCS'05)*, pages 514–524, 2005.

[ZPT04] Xuan Zhou, HweeHwa Pang, and Kian-Lee Tan. Hiding data accesses in steganographic file system. In *Proceedings of the 20th International Conference on Data Engineering*, pages 572–583. IEEE Computer Society, 2004.