Advanced Applications for e-ID Cards in Flanders

ADAPID Deliverable D4

Basic Research

C. Diaz (Ed.), B. De Decker, H. Dekeyser, S. Gevers, M. Layouni, S. Nikova, B. Preneel, X. Sun, C. Troncoso, S. Van Damme, H. Vangheluwe, K. Verslype, and M. Zia

April 2007

Executive Summary

This report presents the basic research done within ADAPID in the first twenty months of the project. The report summarizes the publications, technical reports, and work in progress of the ADAPID partners in several research areas relevant to the project.

First, we present our research on digital credentials, a core technology for secure and privacy-enhanced identity management. Digital credentials are a powerful mechanism that enables secure and flexible identity management protocols, much more advanced than traditional PKI solutions on which the current e-ID is based.

We then present our results in the areas of anonymity metrics and anonymous communications. Anonymity metrics are generic models that serve as tools to evaluate the security of anonymity systems. Anonymous communication channels are a basic building block for privacy-enhanced applications, that enable the protection of the communication layer by preventing traffic analysis attacks.

Our results in the area of secure storage, a basic building block for the project framework, are focused on two main topics: steganographic file systems, which allow users to hide data, and private access to databases, where users are able to query databases without revealing information about the content of their queries.

We also present our research on primitives for distribution of trust. In many scenarios involving the e-ID card it is advisable, for safety reasons, to distribute trust among several servers instead of relying on one single entity. To this end, a number of protocols are used, which allow operations to run correctly even in presence of malicious entities.

Next, our research activities in the area of policies and digital rights management (DRM) is described. Policies define the rules applications will abide with. Also, they make applications more flexible and allow for automation and, hence, a reduction of interaction with the user. DRM allows the owner of private data to control access to and usage of that data.

We present our work on model-based design processes for the engineering of secure and privacy-preserving e-ID applications. Multi-formalism modelling and model transformations are a powerful tool that enable analysis of

required properties of e-ID systems.

In the legal field, four main topics were covered. Firstly, we have looked at issues of data processing and privacy protection regarding the e-ID and advanced e-ID applications in general. Specific attention was given to logging functionality, as well as rules governing analysis of log files and other personal data for scientific research. A second research topic was preservation of evidence, in particular one of the outputs of advanced e-ID applications, namely electronically signed records. The rules of evidence regarding contracts were looked at in particular, since these rules serve as default in other areas. Also, we analysed certain key legal terms and compared them with concepts from the InterPARES framework on digital archiving. Thirdly, the regulatory framework regarding public sector aid was explored. Past experience has shown that the market may not provide the necessary building blocks – or not provide them with the required quality – for the development of advanced e-ID applications. The government may feel the need to enter the market themselves, thus raising questions regarding the regulations governing the financing of public services by public authorities. Last, but not least, the topic of trust and accountability has been the subject of research. Though this research is highly conceptual in nature, it aims to clarify which approach to take regarding shared accountability between systems, users and operators.

List of Contributions

Introduction	COSIC, ICRI, DistriNet and McGill
Digital credentials	McGill, DistriNet and COSIC
Anonymity metrics	COSIC
Anonymous communications	COSIC
Data storage	COSIC
Distribution of trust	COSIC
Policies and DRM	Distrinet
Model-driven approaches	McGill and DistriNet
Legal aspects	ICRI
Conclusions	COSIC, ICRI, DistriNet and McGill
Editor	Claudia Diaz (COSIC)

List of ADAPID Publications

1. Stefan Brands, Liesje Demuynck and Bart De Decker, A practical system for globally revoking the unlinkable pseudonyms of unknown users, *accepted to* Australasian Conference on Information Security and Privacy, 2007.

2. Stefan Brands, Liesje Demuynck and Bart De Decker, A practical system for globally revoking the unlinkable pseudonyms of unknown users, K.U.Leuven, Department of Computer Science, Report CW 472, Leuven, Belgium, December, 2006.

3. Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya, and Mira Meyerovich, *How to win the clonewars: efficient periodic n-times anonymous authentication.*, Proceedings of the 13th ACM Conference on Computer and Communications Security, ACM, 2006, pp. 201–210.

4. George Danezis and Claudia Diaz. *A Survey of Anonymous Communication Channels.* Submitted to the Journal of Privacy Technology, 40 pages.

5. George Danezis and Claudia Diaz. *Improving the Decoding Efficiency of Private Search.*In Anonymous Communication and its Applications, number 05411 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fur Informatik (IBFI), Schloss Dagstuhl, 11 pages, 2006.

6. George Danezis and Claudia Diaz. *Space-Efficient Private Search.* To appear in the Proceedings of the International Conference on Financial Cryptography and Data Security (FC'07), 15 pages, LNCS (in print), Springer, 2007.

7. George Danezis, Claudia Diaz and Carmela Troncoso. *Two-Sided Statistical Disclosure Attack.* Accepted paper at Privacy Enhancing Technologies 2007 (PET'07), 15 pages.

8. Hannelore Dekeyser. *Preservation of Signed Electronic Records.* In Proceedings of the DLM Conference Budapest, 8 p., October 5-7 2005, http://ec.europa.eu/transparency/archival_policy/dlm_forum/proceed2005_en.htm.

9. Hannelore Dekeyser and Jos Dumortier. *Speaking of record's preservation: avoiding terminological confusion between legal experts and archivists.* Submitted to InterPARES (The International Research on Permanent Authentic Records in Electronic Systems, http://www.interpares.org).

10. Hannelore Dekeyser and Sven Van Damme and Jos Dumortier. *Gebruik van informatiesystemen registreren: logbestanden in het licht van C.A.O. nr. 81* (*Logging use of information systems: logfiles in light of C.L.A. nr. 81.*). Submitted to Privacy & Informatie (Privacy & Information journal).

11. Liesje Demuynck and Bart De Decker, How to prove list membership in logarithmic time, K.U.Leuven, Department of Computer Science, Report CW 470, Leuven, Belgium, December, 2006

12. Liesje Demuynck and Bart De Decker, Anonymous updating of credentials, CW report 430, December 2005

13. Liesje Demuynck, Bart De Decker and Wouter Joosen, A Credential-based System for the Anonymous Delegation of Rights, *accepted by* IFIP SEC 2007.

14. Liesje Demuynck and Bart De Decker, Privacy-preserving electronic health records, Communications and Multimedia Security (Dittmann, J. and Katzenbeisser, S. and Uhl, A., eds.), vol 3677, Lecture Notes in Computer Science, pp. 150-159, 2005.

15. Liesje Demuynck and Bart De Decker, Credential-based systems for the anonymous delegation of rights, K.U.Leuven, Department of Computer Science, Report CW 468, Leuven, Belgium, November, 2006

16. Claudia Diaz, Carmela Troncoso and George Danezis. *Does additional information always reduce anonymity?* 10p. Submitted to the Workshop on Privacy in the Electronic Society (WPES'07).

17. Claudia Diaz. *Anonymity Metrics Revisited.* In Anonymous Communication and its Applications, number 05411 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fur Informatik (IBFI), Schloss Dagstuhl, 6 pages, 2006.

18. Claudia Diaz and Bart Preneel. *Accountable Anonymous Communication.* To appear as Chapter of Security, Privacy and Trust in Modern Data Management, 15 pages, Springer.

19. Claudia Diaz, Peter Fairbrother and Carmela Troncoso. *Observable Steganographic File Systems using Pool Mixes and Universal Re-encryption.* 18p. This paper was accepted for publication in April 2007 at the 7th Privacy Enhancing Technologies Workshop.

20. Steven Gevers and Bart De Decker, Privacy friendly information disclosure, On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops (Meersman, R. and Tari, Z. and Herrero, P., eds.), vol 4277, Lecture Notes in Computer Science, pp.636-646, 2006.

21. Steven Gevers, and Bart De Decker, Automating privacy friendly information disclosure, K.U.Leuven, Department of Computer Science, Report CW 441, April, 2006.

22. Steven Gevers and Bart De Decker, Enhancing Privacy in Identity Systems, *submitted to* PET 2007.

23. Emilia Käsper, Ventzislav Nikov, and Svetla Nikova, *Strongly Multiplicative Hierarchical Threshold Secret Sharing*, accepted paper at the International Conference on Information Theoretic Security 2007 (ICITS'07).

24. Mohamed Layouni and Hans Vangheluwe, *Anonymous k-show credentials*, *accepted to* Fourth European PKI Workshop: Theory and Practice, 2007.

25. Sadaf Mustafiz, Ximeng Sun, Jörg Kienzle, Hans Vangheluwe. Model-Driven Assessment of Use Cases for Dependable Systems. In *MoDELS*, pages 558-573, 2006.

26. Sadaf Mustafiz, Ximeng Sun, Jörg Kienzle, Hans Vangheluwe. Model-Driven Requirements Assessment for Dependable Systems. In *Journal on Software & System Modeling*, 2007.

27. Vincent Naessens, Liesje Demuynck, and Bart De Decker, A Fair Anonymous Submission and Review System, Communications and Multimedia Security 2006, (Leitold, H. and Markatos, E., eds.), vol 4237, Lecture Notes in Computer Science, 2006.

28. Vincent Naessens, Liesje Demuynck, and Bart De Decker, A fair anonymous submission and review system, K.U.Leuven, Department of Computer Science, Report CW 442, April, 2006

29. Vincent Naessens, *A methodology for anonymity control in electronic services using credentials*, Phd, Department of Computer Science, K.U.Leuven, Leuven, Belgium, June 2006, ISBN:90-5682-711-1. URL: `http://www.cs.kuleuven.be/publicaties/doctoraten/cw/CW2006\_03.abs.html`

30. Vincent Naessens and Bart De Decker, A methodology for designing controlled anonymous applications, In *Security and Privacy in Dynamic Environments*, volume 201/2006 of *IFIP International Federation for Information Processing*, (Fischer-Hübner, S., Rannenberg, K., Yngström, L. and Lindskog, S., eds.), pages 111–122. IFIP TC11, Springer Boston, 2006.

31. Ventzislav Nikov, Svetla Nikova, *On a Relation between Verifiable Secret Sharing and a Class of Error-Correcting Codes*, In Proceedings of the International Workshop on Coding and Cryptography (WCC 2005), pp. 372-382, 2005, Springer-Verlag, LNCS 3969, 2005, pp. 275-290.

32. Svetla Nikova, Ventzislav Nikov, In Proceedings of the Contact Forum *Coding theory and cryptography*, Royal Flemish Academy for Science and the Arts, Brussels, October 7, 2005, pp. 59-75.

33. Ernesto Posse and Hans Vangheluwe. kiltera: a Simulation Language for Timed, Dynamic-structure Systems. In *Proceedings of The 40th Annual Simulation Symposium*, 2007.

34. Ximeng Sun. A Model-Driven Approach to Scenario-Based Requirements Engineering. 102 pages. Submitted as a Master's thesis (supervised by Prof. Hans Vangheluwe).

35. Ximeng Sun, Hans Vangheluwe. Transforming Software Requirements by Meta-modelling and Graph Transformation. In *Model-Driven Engineering Languages and Systems: 10th International Conference*. MoDELS'07. Submitted.

36. Carmela Troncoso, Claudia Diaz and Bart Preneel. *Traffic Analysis Attacks on a Continuously-Observable Steganographic File System.* This paper was accepted for publication in April 2007 at the 9th Information Hiding Workshop.

37. Sven Van Damme and Reshma Thomas. *Special report on e-government: preserving privacy.* In Global Identification, 42-45, October 2006.

38. Kristof Verslype, Steven Gevers and Bart De Decker, Towards a Privacy-Friendly Next Generation Electronic Identity Card, *submitted to* PET 2007.

39. Kristof Verslype and Bart De Decker, A Flexible and Open DRM Framework, Communications and Multimedia Security 2006, (Leitold, H. and Markatos, E., eds.), vol 4237, Lecture Notes in Computer Science, 2006, pp. 173-184.

40. Miriam Zia, Sadaf Mustafiz, Hans Vangheluwe and Jörg Kienzle. A modelling and simulation based process for dependable systems design. *Journal on Software & System Modeling*, 2007.

41. Miriam Zia, Ernesto Posse, Hans Vangheluwe. Verification of Security Requirements through Multi-formalism Modelling and Model Transformation. In *2nd International Conference on Software and Data Technologies*. ICSOFT 2007. Submitted.

Contents

Chapter 1

Introduction

ADAPID is a project aimed at studying advanced applications for e-ID cards, with a strong focus on security and privacy. Within ADAPID, we are doing research on various technological and legal aspects that would enable these secure and privacy enhanced applications. This report presents a summary of the basic research done in the first year and a half of the project, and includes results on digital credentials, anonymity metrics, anonymous communication, secure storage, trust distribution techniques, policies, modeling and legal issues. The rest of this section motivates and summarizes the main research results that can be found in this report.

With the digitization of society and the migration of services to the electronic world, traditional paper-based credentials have morphed into a digital form. Digital credentials have a number of attractive features that make them superior to their paper counterparts, among which we note searchability, large-scale datamining, and knowledge discovery, just to name a few. With the latter features, came also the disadvantage that credential holders are now a lot easier to monitor, and to have their privacy violated. Furthermore, digital credentials by their very nature are easy to clone and copy, and using them without proper safeguards could lead to serious security problems.

Our research in the area of privacy preserving digital credentials, addresses all the above issues, and others. More specifically, our goal is to develop applications based on the Belgian e-ID card, that preserve the citizens' privacy, as well as the interests of the broader society, including government agencies, businesses, and fellow citizens. The work we do spans a number of application domains such as e-health, e-government, and e-finance.

Anonymity metrics are a generic mechanism to assess the degree of protection provided by anonymity systems to their users. The information theoretic metrics we present in Chapter 3 can be applied to concrete systems, adversaries and conditions. These metrics give a measure of the size and distinguishability of the set of subjects potentially linked to a particular transaction, and attacked by a concrete adversary, able to obtain probabilistic information.

Anonymous communication infrastructures are an essential building block to implement privacy enhanced applications. If the communication layer is not anonymized, then privacy-enhancing techniques applied at the application layer can be rendered ineffective by observations at the communication layer, as users would be identified by their IP addresses. We present our research on anonymous communication in Chapter 4, where we present three papers. The first is a survey of previous research on anonymous communication; the second is a proposal for building an anonymous communication infrastructure that supports re-identification in case of abuse; and the third presents a new attack on anonymous email systems that takes into account replies to messages.

Storage is one of the pillars of the ADAPID framework. In order to develop and support e-health, e-government, and e-business applications, it is necessary a secure archive that preserves electronic documents for long periods of time. Conventional security properties (confidentiality, integrity and availability) need to be fulfilled by a secure archive, beside privacy enhancing techniques that allow its use in applications that imply the use of sensible data (medical data, financial data, etc.).

Our basic research has focused on one side on secure storage with no leak of information, in the form of Steganographic file systems (see Sect. 5.2 and 5.3), where a user or an institution can store and access data in a secure and privacy-enhanced way. A second line of research studied the field of private search, which enhance the private access to databases (Sect. 5.4), concealing the exact information searched even for the database holder.

In chapter 6 our research on protocols for distribution of trust is presented. To distribute trust among a group of entities is desirable in many scenarios. The goal is certain specified groups of them to be able to perform an operation, but smaller, malicious subsets of them not to be able to do any harm. To this end protocols like secret sharing schemes, verifiable secret sharing schemes, and multi-party computation are used.

We have performed research on some of these protocols. We have focused on some relations between verifiable secret sharing and a class of error-correcting codes (see Sect. 6.2) and on multi-party computation protocols in a hierarchical setting, where participants have different capabilities depending on their position in the hierarchy (see Sect. 6.4.) We have also made an overview on secrets sharing schemes, which is presented in Sect. 6.3.

Policies are important in privacy friendly applications. They define the rules by which the applications will abide. They serve different purposes: they may be used to convey these rules to another application (e.g. a client), they allow for customization of an application or can reduce user interaction. A distinction is made between privacy, disclosure and access policies. Privacy policies are used by service providers to describe the personal information that a user will have to provide in order to get access to a particular service. They also specify the purpose of that data, its processing, retention and possible disclosure to third parties. Disclosure policies enable users to restrict the disclosure of personal data (or properties thereof) to service providers. They also automate in most cases the selection of the proper credentials to use for a particular service. Access policies regulate the access to and the usage of resources. They form the basis for authorization decisions to be made in the applications.

DRM of Digital Rights Management is the generic term for techniques used by publishers or copyright owners to control access to and usage of digital data or hardware. It may also refer to the restrictions associated with a specific instance of a digital work or device.

The research done in this area is twofold: on the one hand, existing privacy and disclosure policy languages are extended to support the use of private credentials and the selection of the proper credentials is being automated; on the other hand, a secure and flexible DRM system is designed, and the use of DRM techniques to protect personal data and enforce policies are being studied.

Complex systems need to address a variety of requirements early on in the development process. To tackle issues related to satisfying these requirements, model-based approaches are increasingly used in all stages of the system design. Multiple models are introduced to expose a different view on the application, and transformation rules are defined to generate a different view on the application automatically. These multi-formalism modeling methodologies control the complexity of the design process, and allow us to perform analysis on the applications even before they are developed. This analysis is used to validate whether the designed e-ID application respects the required anonymity, privacy, and security constraints. Various model-based approaches are proposed in Chapter 8 to address different e-ID application requirements.

The aim of the legal chapter is to prevent or reduce friction between technology and the legal framework, thus increasing the chances of a successful deployment. Issues of privacy protection are a major source of friction, which need to be addressed from the design phase. Other cause for concern is accountability, not only of the advanced e-ID application operator towards the users, but also between users. Specifically, the question of availability of credible evidence arises. When it comes to deployment of advanced e-ID applications, market-failure is a risk to be considered. Without going into an economical analysis, we look at the possibilities for governments to step in.

Chapter 2

Privacy-Preserving Digital Credentials

2.1 Introduction

In a transaction-based world, with continuously shrinking resources, it is no surprise that access control has always been, and still continue to be a central issue. Often times, to benefit from a service or a resource, one is asked to pay a fee or prove possession of a set of qualifications and privileges. In general, such proofs and payments are made using various types of credentials. Over time, the notion of "credentials" has evolved in a quite impressive way. At the very beginning, a credential used to be an oral or common-knowledge assertion by a relative or a central authority in the clan about the identity of a certain clan member. With the development and expansion of human societies, it became clear that the previous notion of credentials was not reliable any more, and credentials were replaced by paper-based documents issued by commonly-recognized central authorities. Examples of widely used credentials include national identity cards, property certificates, bank notes etc. Although, proofs with the new credentials were much easier to establish and verify, the latter still had a number of shortcomings. For instance, despite the certification authorities' best efforts, it is still feasible, and in some cases relatively easy, to counterfeit credential documents. In addition, with paper-based credentials, it is sometimes hard to prevent users from lending their personal credentials to a third party, or from pooling their credentials together in order to access a service to which they are not entitled individually. Furthermore, paper-based credentials can be used to infringe on their holders' liberties, and may expose them to various forms of coercion, extortion, and robbery. Another major limitation of paper-based credentials, is their lack of privacy for their holders; for instance, if accessing a certain service requires applicants to be *citizens* of a given country, then using paper-based credentials, a credential holder will leak a lot more information than his or her citizenship. This unnecessary flow of extra information is often undesirable and can be used in various forms of discrimination and profiling.

With the digitization of society and the migration of services to the electronic world, traditional paper-based credentials have morphed into a digital form. Digital credentials have a number of attractive features that make them superior to their paper counterparts, among which we note searchability, large-scale data-mining, and knowledge discovery, just to name a few. With the latter features, came also the disadvantage that credential holders are now a lot easier to monitor, and to have their privacy violated. For instance, in 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 infinitely-many 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, UPr, Hig].

Relevance to ADAPID

One of ADAPID's mandates is to develop applications based on the Belgian e-ID card, that preserve the citizens' privacy, as well as the interests of the broader society, including government agencies, businesses, and fellow citizens. Ideally, the e-ID card should ensure both the privacy and accountability properties specified in the previous section. The research we are doing in the area of privacy-preserving digital credentials is directed exactly towards achieving the above mentioned goals. More specifically, we are studying various extensions, and novel applications of existing credential systems, in contexts such as e-health, e-government, and e-business. In the remainder of this chapter we present a number of published and submitted research papers on privacy preserving digital credentials. For instance, in Section 2.3 we present an efficient single sign-on system in which a user can access services using unlinkable digital pseudonyms that can all be revoked in case he or she abuses any one service. In Section 2.10, we describe a construction to achieve anonymous k-show credentials. Anonymous k-show credentials allow users to authenticate up to $k > 1$ times, without their identity being revealed. The user's identity is unveiled only if the credential is shown more than the allowed k times. In Section 2.11, we present a credential system that allows a user to anonymously authenticate at most n times, in a single time period. The presented constructions have a number of application domains such as e-government, e-health etc. In the following we highlight few of them:

- Payment of Social Benefits: People entitled to receive social benefits or discounts up to a certain limit could benefit from our techniques. The credential holder would be able to anonymously request the benefits he or she is entitled to, e.g., once a month. However, in case of abuse the system would deanonymize misbehaving users (e.g., those who try to get benefits twice in a month).

- Prescriptions for chronical diseases: Patients who suffer from chronical diseases and need permanent treatment could have such a credential for buying up to n medicine boxes per month. Every month the ability to buy the medicines would be renewed, so that patients would remain anonymous as long as they do not exceed the amount of medicines they are buying. This mechanism combines privacy properties with abuse prevention (e.g., people trying to buy large quantities of medicines for re-selling would be identified).

- Public-transit passes: Limited show credentials can be used to issue public transit passes, where a user is allowed to make up to k rides anonymously, after which the pass's serial number will be uncovered, revoked, and added to a black list.

2.2 A Credential-based System for the Anonymous Delegation of Rights

- Liesje Demuynck, Bart De Decker and Wouter Joosen, A Credential-based System for the Anonymous Delegation of Rights, *accepted by* IFIP SEC 2007.

- Liesje Demuynck and Bart De Decker, Credential-based systems for the anonymous delegation of rights, K.U.Leuven, Department of Computer Science, Report CW 468, Leuven, Belgium, November, 2006

Summary

An anonymous delegation system enables individuals to retrieve rights and to delegate different subparts of these rights to different entities. The delegation procedure is anonymous, such that no collusion of entities can track an individual's delegation behavior. On the other hand, it is ensured that a user cannot abuse her delegation capabilities. This report introduces a general delegation model and presents three schemes. The schemes are based on credential systems and provide both anonymity for the individual and security for the organizations. The schemes are compared based on their functionality, privacy and security characteristics. Additionally, some guidelines are given for choosing a particular implementation based on the application's requirements.

The published paper presents the most sophisticated of the three schemes.

Contributions

Three different delegation schemes are described. They differ in functionality (expressiveness of the constraints on usage and on the delegation process), in privacy properties (unlinkability and indistinguishability), in monitorability (immediate or delayed) and in complexity. Depending on the required properties, the most appropriate scheme can be chosen. Note that delegation is never absolute. In each scheme, a revocation protocol is provided.

2.3 A practical system for globally revoking the unlinkable pseudonyms of unknown users

- Stefan Brands, Liesje Demuynck and Bart De Decker, A practical system for globally revoking the unlinkable pseudonyms of unknown users, *accepted to* Australasian Conference on Information Security and Privacy, 2007.
- Stefan Brands, Liesje Demuynck and Bart De Decker, A practical system for globally revoking the unlinkable pseudonyms of unknown users, K.U.Leuven, Department of Computer Science, Report CW 472, Leuven, Belgium, December, 2006.

Summary

In the report and the paper, we propose the first single sign-on system in which a user can access services using unlinkable digital pseudonyms that can all be revoked in case he or she abuses any one service. Our solution does not rely on key escrow: each user needs to trust only her own computing device with following our protocols in order to be assured of the unconditional untraceability and unlinkability of her pseudonyms. In applications where users hook pseudonyms up to legacy identifiers or legacy accounts at service providers, our system guarantees that service providers (even if they collude with the pseudonym issuer) do not gain any correlation powers over users. Our solution is highly practical.

Contributions

The proposed solution involves two novel ingredients:

- a technique for invisibly chaining all of a user's pseudonyms in a manner that permits the revocation of all of them on the basis of any one of them (without knowing the user's identity with the issuer)
- and a sublinear-time proof that a committed number is not on a blacklist without revealing additional information about the number.

2.4 How to prove list membership in logarithmic time

- Liesje Demuynck and Bart De Decker, How to prove list membership in logarithmic time, K.U.Leuven, Department of Computer Science, Report CW 470, Leuven, Belgium, December, 2006

Summary

In this report we propose a novel technique for proving in zero-knowledge that a committed value is on a public list. Our technique combines the concept of a hashtree with well-known zero-knowledge proofs of knowledge. The resulting proof protocol has a time complexity logarithmic in the size of the list. Its soundness property is guaranteed under the discrete logarithm assumption, while the prover's secret information is protected even in the face of a computational unbounded adversary. Applications of our technique can be found in membership revocation for group signature schemes, anonymous credential systems and identity escrow schemes.

Contributions

The major contribution of this report is the design of a protocol by adapting and combining existing techniques in order to achieve the resulting complexity gain, while still being secure under the well-accepted discrete logarithm (DL) assumption. Moreover, the protocol guarantees unconditional privacy even against a computationally unbounded verifier.

2.5 Anonymous updating of credentials

- Liesje Demuynck and Bart De Decker, Anonymous updating of credentials, CW report 430, December 2005

Summary

In this report, a high-level overview of credential systems is presented. Particular attention is paid to the two most elaborated systems to date: the credential system by Brands on the one hand and the CL system of Camenisch and Lysyanskaya on the other hand. Both systems are described in detail and a comparison is made based on their cryptographic building blocks, linkability properties and performance measures. It is then shown how the protocols in both systems can be extended to enable the anonymous updating of credentials. The resulting systems are evaluated using a case study implementing a privacy-friendly loyalty card.

Contributions

The report describes two protocols (one for Idemix, another for Brands' credentials), which allow for anonymously updating of credentials. The update may involve one or more:

- additions of a delta value to an attribute (while the issuer does not know the original value),
- assignments of a new value to an attribute (the value can be chosen by the issuer, the credential holder, or both)

The new credential is completely unlinkable to the original one.

2.6 Enhancing Privacy in Identity Systems

- Steven Gevers and Bart De Decker, Enhancing Privacy in Identity Systems, *submitted to* PET 2007.

Summary

On the internet today, users disclose a lot of information. To help these users to protect their privacy, user centric identity management systems (IMS) put them in control of their personal data. Users know exactly what information they disclose to organizations. Private certificate systems, however, offer some capabilities that enhance the privacy of the user that are not yet included in IMS. For example, by using private certificates, it is possible to prove properties of claims (e.g. being older than eighteen). This paper introduces privacy enhanced claim URIs that make it possible to request information of a user in a privacy friendly way. Then, it is shown how these URIs allow integration of private certificate systems in Microsoft Cardspace. Also, many capabilities of private certificates can be achieved in current IMS, even without using private certificates. The user's privacy is better protected, while very little changes have to be made to the IMS. Since our approach is very simple and widely applicable, it allows to make many of today's online transactions much more privacy friendly.

Contributions

The paper proposes a way to specify privacy enhanced claims using URIs. Using these URIs, many identity management systems can provide better privacy properties. Also, the URIs allow to integrate private certificates in Microsoft Cardspace.

2.7 Towards a Privacy-Friendly Next Generation Electronic Identity Card

- Kristof Verslype, Steven Gevers and Bart De Decker, Towards a Privacy-Friendly Next Generation Electronic Identity Card, *submitted to* PET 2007.

Summary

Electronic identity (e-ID) cards are introduced increasingly. Digital credentials are gaining importance as well. This poses a threat to the privacy of the owner since e-ID cards and digital credentials generally contain a unique identier of the owner. Moreover, a user will have to manage dozens or hundreds of credentials in the future: a credential for his sporting club, a digital driving license, e-tickets for the cinema, etc. The owner will want to use these credentials wherever and whenever he/she wants. The credentials must remain managable as well and, in case of theft or loss, they must become unusable by others and recoverable by the legitimate owner. This paper proposes a next-generation e-ID card, which tries to solve the above issues by using the privacy-friendly anonymous credentials.

Contributions

The paper presents a secure and privacy friendly way to manage anonymous credentials using an online repository (server) which avoids loss of credentials. Previous credential repositories required a higher level of trust in the server storing the credentials and credentials could be read by the server or could be linked to the same user. Our solution solves these weaknesses and prevents tampering by the server as well. A personal smart card (in fact, an e-ID card) is required to access and use the credentials. A privacy, security, storage and time analysis was made. The conclusion is that currently, still a potentially untrusted client is needed to run the main part of the protocols.

2.8 A Fair Anonymous Submission and Review System

- Vincent Naessens, Liesje Demuynck, and Bart De Decker, A Fair Anonymous Submission and Review System, Communications and Multimedia Security 2006, (Leitold, H. and Markatos, E., eds.), vol 4237, Lecture Notes in Computer Science, 2006.

- V. Naessens, L. Demuynck, and B. De Decker, A fair anonymous submission and review system, K.U.Leuven, Department of Computer

Science, Report CW 442, April, 2006

Summary

Reputation systems play an important role in many Internet communities. They allow individuals to estimate other individual's behavior during interactions. However, a more privacy-friendly reputation system is desirable while maintaining its trustworthiness. This paper presents a fair anonymous submission and review system. The review process is reputation-based and provides better anonymity properties than existing reputation systems. Moreover, the system allows for accountability measures. Anonymous credentials are used as basic blocks.

Contributions

This paper presents a fair anonymous submission and review system.

- The system provides a trustworthy environment for authors, reviewers and conference chairmen. Authors and reviewers are anonymous. However, double submissions of papers can be dealt with.
- The review process is reputation-based and allows for accountability measures (e.g. inferior reviews). The reputation is based on prior publications.
- The reputation system makes use of updatable anonymous credentials.

2.9 Privacy-preserving electronic health records

- Liesje Demuynck and Bart De Decker, Privacy-preserving electronic health records, Communications and Multimedia Security (Dittmann, J. and Katzenbeisser, S. and Uhl, A., eds.), vol 3677, Lecture Notes in Computer Science, pp. 150-159, 2005.

Summary

Electronic health records enable the global availability of medical data. This has numerous benefits for the quality of offered services. However, privacy concerns may arise as now both the patient's medical history as well as the doctor's activities can be tracked. In this paper, we propose an electronic health record system which allows the patient to control who has access to her health records. Furthermore, provided she does not misuse the system, a doctor will remain anonymous with respect to any central authority.

Contributions

In this paper we have described a secure and privacy-preserving electronic health record system.

First, a technique is presented to link invisibly all the patient's e-health records, by hashing a secret value, only known to the patient and the doctor in attendance. The secret value can be modified, which allows for revoking a doctor's access to the the e-health records. Moreover, a doctor can easily derive all the previous secret values from the current one. Hence, the patient keep e-health records hidden from her doctor.

Secondly, several protocols are designed for visiting a doctor (and, hence, giving access to the EHRs), retrieving and adding new EHRs (by proving knowledge of the current secret), changing the secret key and performing an emergency retrieval.

2.10 Anonymous k-show Credentials

- Mohamed Layouni and Hans Vangheluwe, *Anonymous k-show credentials*, *accepted to* Fourth European PKI Workshop: Theory and Practice, 2007.

Summary

The paper proposes an extension of the Idemix [CL01a] credential system to k-show credentials, for $k > 1$. Limited-show credentials are a useful tool for a variety of applications. For instance, they can be used to build public transit passes, where a user is allowed to make up to k rides anonymously, after which the pass's serial number will be uncovered, revoked, and added to a black list. In order to count the number of times a credential is shown, the issuing organization is able to link different showings of the same credential to each others, but not to the identity of the credential holder, or for that matter, to the instance of the issuing protocol that generated the credential. This linking feature can also be found in the one-show credentials of [CL01a] and [Bra94] where issuers rely on it to detect double-spending.

Contributions

The Idemix credential system [CL01a, CL04a] offers only two types of credentials:

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

2. One-show credentials: they can be shown anonymously only once, before the identity of their holder is unveiled.

In this paper, we bridge the gap between the two first types of credentials, and extend the Idemix framework to k-show credentials (for $k > 1$.), which can be shown anonymously up to k times, after which the identity of the holder is revealed. A naive way to construct k-show credentials is by issuing k separate copies of one-show credentials, but this option obviously lacks efficiency. The solution we propose in this paper extends the one-time show credentials of [CL01a] to k-show credentials without a significant increase in complexity. Compared to the protocols of [CL01a], we only add 2 extra exponentiations and 1 proof of DL knowledge to the user in the pseudonym creation protocol. For the issuing protocol, the user performs 3 more exponentiations and a proof of knowledge for each additional showing allowed. Finally, the complexity of the showing protocol can be made very close to that of one-time show credentials [CL01a] by using precomputations and fast exponentiation methods [Gor98].

2.11 Efficient Periodic n-Times Anonymous Authentication

- Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya, and Mira Meyerovich, *How to win the clonewars: efficient periodic n-times anonymous authentication.*, Proceedings of the 13th ACM Conference on Computer and Communications Security, ACM, 2006, pp. 201–210.

Summary

We create a credential system that lets a user anonymously authenticate at most n times in a single time period. A user withdraws a dispenser of n e-tokens. She shows an e-token to a verifier to authenticate herself; each e-token can be used only once, however, the dispenser automatically refreshes every time period. The only prior solution to this problem, by Damgard et al. in [DDP06], uses protocols that are a factor of k slower for the user and verifier, where k is the security parameter. Damgard et al. also only support one authentication per time period, while we support n. Because our construction is based on e-cash, we can use existing techniques to identify a cheating user, trace all of her e-tokens, and revoke her dispensers. We also offer a new anonymity service: glitch protection for basically honest users who (occasionally) reuse e-tokens. The verifier can always recognize a reused e-token; however, we preserve the anonymity of users who do not reuse e-tokens too often.

Background

As computer devices get smaller and less intrusive, it becomes possible to place them everywhere and use them to collect information about their environment. For example, with todays technology, sensors mounted on vehicles may report to a central traffic service which parts of the roads are treacherous, thus assisting people in planning their commutes. Yet this vision appears to be incompatible with privacy. A sensor installed in a particular car will divulge that cars location.

A naive solution would be to supply only the relevant information and nothing else. A report about the road conditions should not say which sensor made the measurement. However, then nothing would stop a malicious party from supplying lots of false and misleading data. We need to authenticate the information reported by a sensor without divulging the sensors identity. We also need a way to deal with rogue sensors, i.e., formerly honest sensors with valid cryptographic keys that are captured by a malicious adversary and used to send lots of misleading data

A suite of cryptographic primitives such as group signatures [ACJT00, BBS04, CS97, CvH91] and anonymous credentials [CL01b, CL02, CL04b, Cha85b, Dam90, LRSW99] has been developed to let us prove that a piece of data comes from an authorized source without revealing the identity of that particular source. However, no previous results provide a way to ensure anonymity and unlinkability of honest participants while at the same time guaranteeing that a rogue cannot undetectably provide misleading data in bulk. Indeed, it seems that the ability to provide false data is a consequence of anonymity.

Recently Damgard, Dupont and Pedersen presented a scheme in [DDP06] that overcomes this seeming paradox. The goal is to allow an honest participant to anonymously and unlinkably submit data at a small rate (for example, reporting on road conditions once every fifteen minutes), and at the same time to have a way to identify participants that submit data more frequently. This limits the amount of false information a rogue sensor can provide.

While the work of Damgard et al. is the first step in the right direction, their approach yields a prohibitively expensive solution. To authenticate itself, a sensor acts as a prover in a zero-knowledge (ZK) proof of knowledge of a relevant certificate. In their construction, the zero-knowledge property crucially depends on the fact that the prover must make some random choices; should the prover ever re-use the random choices he made, the provers secrets can be efficiently computed from the two transcripts. The sensors random choices are a pseudorandom function of the current time period (which must be proven in an additional ZK proof protocol). If a rogue sensor tries to submit more data in the same time period, he will have to use the same randomness in the proof, thus exposing his identity. It is very challenging to instantiate this solution with efficient building blocks. Damgard et al. use the most efficient building blocks available, and also introduce some of their own; their scheme requires that the user perform $57 + 68k$ exponentiations to authenticate, where k is the security parameter (a sensor can cheat with probability $2 - k$).

We provide a completely different approach that yields a practical, efficient, and provably secure solution. We relate the problem to electronic cash (e-cash) [Cha82, Cha83] and in particular, to compact e-cash [CHL05]. In our approach, each participant obtains a set of e-tokens from the central server. Similar to the withdrawal protocol of e-cash, the protocol through which a participant obtains these e-tokens does not reveal any information to the server about what these e-tokens actually look like. Our protocol lets a participant obtain all the e-tokens it will ever need in its lifetime in one efficient transaction. The user performs only 3 multi-base exponentiations to obtain e-tokens, and 35 multi-base exponentiations to show a single e-token. If the user is limited to one e-token per time period (as in the Damgard et al.s scheme), the scheme can be further simplified and the user will need to do only 13 multi-base exponentiations to show an e-token.

Distributed sensors can use an e-token to anonymously authenticate the data they send to the central server. In the on-line game scenario, each e-token can be used to establish a new connection to the game. Unlike e-cash, where it is crucial to limit the amount of money withdrawn in each transaction, the number of e-tokens obtained by a participant is unlimited, and a participant can go on sending data or connecting to the game for as long as it needs. The e-tokens are anonymous and unlinkable to each other and to the protocol where they were obtained. However, the number of e-tokens that are valid during a particular time period is limited. Similarly to what happens in compact e-cash, reusing e-tokens leads to the identification of the rogue participant. We also show how to reveal all of its past and future transactions.

Thus, in the sensor scenario, a sensor cannot send more than a small number of data items per time period, so there is a limit to the amount of misleading data that a rogue sensor can submit. Should a rogue sensor attempt to do more, it will have to reuse some of its e-tokens, which will lead to the identification of itself and possibly all of its past and future transactions.

Contributions

Our main contribution is the new approach to the problem, described above, that is an order of magnitude more efficient than previous solutions. We present a basic construction, which is based on previously-proposed complexity theoretic assumptions (SRSA and y-DDHI) and is secure in the plain model. Our construction builds on prior work on anonymous credentials, so that it is easy to see which parts need to be slightly modified, using standard techniques, to add additional features such as an anonymity revoking trustee, identity attributes, etc. The computational cost of these additional features is a few additional modular exponentiations per transaction.

We then extend our basic solution to make it tolerate occasional glitches without disastrous consequences to the anonymity of a participant. Suppose that a sensor gets reset and does not realize that it has already sent in a measurement. This should not necessarily invalidate all of the sensors data. It is sufficient for the data collection center to notice that it received two measurements from the same sensor, and act accordingly. It is, of course, desirable, that a sensor that has too many such glitches be discovered and replaced. Our solution allows us to be flexible in this respect, and tolerates m such glitches (where m is specified ahead of time as a system-wide parameter) at the additive cost of $O(km)$ in both efficiency and storage, where k is the security parameter. This does not add any extra computational or set-up assumptions to our basic scheme.

We consider additional variations of our basic scheme. We show, also in the plain model, how to enable the issuer and verifiers to prove to third parties that a particular user has (excessively) reused e-tokens (this is called weak exculpability); and enable the issuer and verifiers to trace all e-tokens from the same dispenser as the one that was excessively reused (this is called tracing). We also show, in the common-parameters and random-oracle models, how to achieve strong exculpability, where the honest verifiers can prove to third parties that a user reused a particular e-token. Finally, we explain how e-token dispensers can be revoked; this requires a model where the revocation authority can continuously update the issuers public key.

2.12 Future work

The area of privacy-preserving digital credentials is a wide and challenging area of research, with a multitude of interesting questions that still need to be answered. We are currently working on a number of issues, among which we note the development of privacy-enhanced databases to guard against non-consensual disclosures of personal data from central databases (McGill). We are also conducting an in-depth comparative survey of the state of the art in privacy preserving digital credentials (joint work DistriNet and McGill).

Chapter 3

Anonymity metrics

3.1 Introduction

The need of a metric to measure the performance of anonymity implementations appeared with the development of applications that enabled anonymous electronic transactions, such as untraceable email, electronic voting, anonymous e-coins or privacy-enhanced web browsing. The research questions that arose were: how can anonymity be measured? How can two different anonymity systems be compared? Is there a general anonymity metric which can be applied to any anonymity system? How can we evaluate the effectiveness of different attacks on the anonymity system? How can we quantify losses and gains in anonymity? How can anonymity metrics reflect the partial or statistic information often obtained by an adversary? Information theoretic anonymity metrics provide answers to these questions.

Information theoretic anonymity metrics can be applied to concrete systems, adversaries and conditions. These metrics give a measure of the size and distinguishability of the set of subjects potentially linked to a particular transaction, and attacked by a concrete adversary. In order to get an idea on the performance of an anonymity implementation under different conditions, multiple anonymity measurements must be made and analyzed.

Information theoretic metrics can be applied to a broad range of anonymity systems. It is thus important to understand the concepts behind entropy-based anonymity metrics in order to apply and interpret them correctly in concrete scenarios. The metrics must be adapted to the anonymity system under study, and the computation of probability distributions that lead to meaningful metric values is not always obvious. In this chapter, we present two papers in the area of anonymity metrics.

Relevance to ADAPID

Anonymity metrics are an essential tool to evaluate the anonymity properties of a given system. They also enable objective comparison of different approaches to provide anonymity services.

In this respect, our results have generic applications, and can be used in the framework of ADAPID to evaluate the degree of privacy protection provided by the used technologies.

3.2 Information theoretic anonymity metrics

- Claudia Diaz. *Anonymity Metrics Revisited.* In Anonymous Communication and its Applications, number 05411 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fur Informatik (IBFI), Schloss Dagstuhl, 6 pages, 2006.

Summary

In 2001, two information theoretic anonymity metrics were proposed: the effective anonymity set size [SD02] and the degree of anonymity [DSCP02]. We propose in [Dia06] an abstract model for a general anonymity system which is consistent with the definition of anonymity on which the metrics are based. We revisit entropy-based anonymity metrics, and we apply them to Crowds, a practical anonymity system. We discuss the differences between the two metrics and the results obtained in the Crowds example.

Background

Prior to the quantification of anonymity, a working definition for the term anonymity was needed. Pfitzmann and Hansen [PH01] defined anonymity as the state of being not identifiable within a set of subjects, the anonymity set. This definition, first proposed in year 2000, has been adopted in most of the anonymity literature.

According to the Pfitzmann-Hansen definition of anonymity, the subjects who may be related to an anonymous transaction constitute the anonymity set for that particular transaction. A subject carries on the transaction anonymously if he cannot be distinguished (by an adversary) from other subjects. This definition of anonymity captures the probabilistic information obtained by adversaries trying to identify anonymous subjects. Information theoretic anonymity metrics aim at giving a meaningful measure of the anonymity set size.

The information theoretic concept of entropy [Sha48] provides a measure of the uncertainty of a random variable. Entropy-based anonymity metrics give a measure of the uncertainty of the adversary on the subject who is related to a transaction of interest. The effective anonymity set size takes into account the number of potential subjects linked to the transaction, and the probabilities assigned to the subjects.

The metric (and thus anonymity) increases its value with two factors. First, with the number of subjects potentially linked to the transaction; and second, with the uniformity of the probability distribution. The more equally distributed the probabilities assigned to the subjects of the anonymity set, the higher the entropy (i.e., the higher the effective anonymity set size).

Contributions

We propose a general model for anonymity systems and present two flavors of information theoretic metrics. With these metrics we can quantify the effective anonymity set size and the degree of anonymity provided by a system in particular attack circumstances. We have applied the metrics to Crowds, an existing solution for anonymous communication, and discussed the results obtained.

Anonymity metrics provide relevant information on the anonymity of concrete subjects in concrete attack scenarios. In order to know more about the robustness of an anonymity system, we need to make multiple measurements in different scenarios.

3.3 On anonymity metrics and conditional entropy

- Claudia Diaz, Carmela Troncoso and George Danezis. *Does additional information always reduce anonymity?* 10p. Submitted to the Workshop on Privacy in the Electronic Society (WPES'07).

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 is more uncertain on the identity of the subject behind an action as there are more subjects in the anonymity set. But his uncertainty also depends 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 information theoretic concept of *Shannon entropy* [Sha48] (or simply "entropy") is a measure of the uncertainty associated with a random variable. Technical measures of anonymity [DSCP02, SD02] are based on the entropy of the probability distribution linking an action to all possible subjects who may be related to it. The anonymity metric gives, thus, a measure of the uncertainty of the attacker.

Shannon entropy has been very useful in the evaluation [DP04, DSD04, DS03] of mix-based [Cha81] 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.

It has been claimed that an adversary with access to more information is *always* able to reduce anonymity [CS06]. However, we show in this paper 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.

The key misunderstanding stems from the confusion of the attacker's uncertainty in a given scenario with Shannon's *conditional entropy* [Sha48]. We explain here that the attacker's uncertainty is given by the entropy of a conditional probability distribution: the probability that a message was sent to each possible recipient, given a user sending profile and a concrete observation of the communication layer. The entropy of this probability distribution is ***not*** the *conditional entropy*. Therefore, known properties of the *conditional entropy* are not applicable to the attacker's uncertainty.

Contributions

Shannon entropy has proven to be very useful to measure anonymity, both in the application (user profiles) and communication layers. Often, entropy based anonymity metrics have been used for the evaluation of mixes, without taking into account any user profile information the attacker could have. Integrating several sources of information has not yet been fully addressed.

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 have explained 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).

Chapter 4

Privacy-Enhanced Infrastructure: Anonymous Communications

4.1 Introduction

A lot of traditional computer security has focused on protecting the content of communications by insuring confidentiality, integrity or availability. Yet the meta data (*traffic data*) associated with it also contains important information in itself. Traffic data comprises the time and duration of a communication, the detailed shape of the communication streams, the identities of the parties communicating, and their location. The knowledge of what typical communication patterns might look like can also be used to infer information about an observed communication; e.g., it can be used to quickly select targets for further surveillance, or to extract information about communications content.

Communication infrastructures, on which state and economic actors are increasingly reliant, are more and more vulnerable to such attacks: wireless and GSM telephony are replacing traditional systems, routing is transparent and protocols are overlayed over others giving the attackers plenty of opportunity to observe, and take advantage of such traffic data. Attackers can make use of this information to gather strategic information, or to attack security protocols and violate traditional security properties. Traffic analysis techniques can naturally be applied to Internet communications, as the access to routers and communication links can be gained by adversaries.

People increasingly use the Internet for an ever wider range of activities: reading the newspaper, shopping, staying in contact with family and friends, finding a partner, booking holidays, expressing their opinion, keeping an online diary, etc. While performing an online activity, even if the confidentiality of the information being transmitted is protected through encryption, the source and destination of the communication are easily traceable. The information on who communicates with whom may reveal critical information that could be used against the Internet user. For example, someone accessing a web site with information on a life-threatening disease may not obtain a health insurance or lose his job if this information gets to the insurance company or the employer.

The linkability of all traffic information generated by an Internet user (e.g., through the IP address, national ID number or social security number), allows for sophisticated profiling of each user. Some of the data that could be gathered and stored directly or indirectly, just by monitoring the users communication are: email address, age, gender, location, religious preferences, sexual orientation, bank, job, type of products bought on the Internet, period of holidays, political orientation, lifestyle, or social network.

In the current communication infrastructure, traffic data is available at moderate cost to anyone willing to harvest it, without the data subject being aware of it. There is already an emerging market of personal data that criminals use to impersonate their victims. In some cases, the damage inflicted to identity theft victims is huge. With the development of data collection technologies, data storage capacity and profiling techniques, these data become easier to get, cheaper to store and more profitable to use (either for legal or illegal practices). Software tools which can be used for privacy violations are increasingly available. These include spyware such as key loggers (e.g., KeyLog Pro [Pro]) or search engines like Google [Goo].

In this scenario, large amounts of information about large numbers of people are under the control of a few data holders. Users effectively lose control over their own personal data, and at the same time all their online activity can be collected, aggregated and stored. These data can be used to take decisions that impact the data subjects. This asymmetric distribution of information creates dangerous imbalances, as those in control of the information may use their acquired power for many different purposes.

The European Directive on Data Protection [PC95] is an attempt to protect citizens from these threats. However, it is very difficult to enforce, and its practical effectiveness is yet to be proven. Technology can be designed to keep personal data under the control of the user. The user could disclose the minimal amount of information to the entities with which he interacts.

Anonymity technologies serve as tools for the protection of privacy in electronic applications, and they are a key component of Privacy Enhancing Technologies (PETs). Anonymous communication networks protect the privacy of Internet users towards the other end of the communication and towards observers in the network. This is achieved by hiding the link between the initiator of the communication and the responder. For applications such as electronic voting and electronic payments, anonymity and privacy are strictly necessary. In a democratic society, public elections will be held anonymously and citizens have a fundamental right to privacy, for example when buying goods or subscribing to services.

Anonymous communication channels are an essential building block of privacy-enhanced systems. Without protection at the communication layer, users can be identified, no matter what sophisticated identity management techniques are implemented. Often, PETs assume the existence of anonymous channels that protect the communication layer, but in fact, low-latency anonymous channels resistant to powerful traffic analysis attacks do not exist. Existing systems are vulnerable to traffic correlation attacks. Better levels of protection can be achieved for low-latency applications. The basic technique for low-latency anonymous communication is the mix, and its anonymity properties are well understood. Nevertheless, mixes are still vulnerable to a series of attacks.

Anonymous communications, despite being first proposed over 25 years ago, has only since 2000 become an extremely active field of research. It is also increasingly relevant since systems that are the direct result of this research, like Tor [DMS04], JAP [JAP] and Mixminion [DDM03], are being deployed and used to protect the privacy of thousands of people.

The ADAPID research on anonymous communications presented in the rest of this chapter includes:

- A survey of anonymous communication techniques, proposals and vulnerabilities.

- A proposal for building anonymous communication infrastructures that combine privacy protection with accountability techniques for misbehaving users.

- A new attack on high-latency, mix-based anonymous email system that takes into account the effects of message replies.

Relevance to ADAPID

Anonymous communication channels are a basic building block for any privacy-enhanced application: protection at the application layer may be rendered useless if the adversary can deploy traffic analysis attacks. For example, the use of sophisticated identity management systems would not protect users' privacy if they can be identified, tracked and profiled though their IP addresses.

In this respect, research on anonymous communications is key in order to design and deploy secure and privacy-enhanced applications, given that all privacy-enhanced ADAPID applications need protection at the communication layer.

4.2 Survey on anonymous communication

- George Danezis and Claudia Diaz. *A Survey of Anonymous Communication Channels.* Submitted to the Journal of Privacy Technology, 40 pages.

Summary

We present in [DD07b] an overview of the field of anonymous communications, from its establishment in 1981 from David Chaum to today. Key systems are presented categorized according to their underlying principles: semi-trusted relays, mix systems, remailers, onion routing, and systems to provide robust mixing. We include extended discussions of the threat models and usage models that different schemes provide, and the trade-offs between the security properties offered and the communication characteristics different systems support.

Background

Research on anonymous communications started in 1981 with Chaum's seminal paper "Untraceable electronic mail, return addresses, and digital pseudonyms" [Cha81]. Since then, a body of research has concentrated on building, analyzing and attacking anonymous communication systems. In this survey we look at the definition of anonymous communications and the major anonymous communication systems grouped in families according to the key design decisions they are based on.

Data communication networks use addresses to perform routing which are, as a rule, visible to anyone observing the network. Often addresses (such as IP addresses, or Ethernet MACs) are a unique identifier which appear in all communication of a user, linking of all the user's transactions. Furthermore these persistent addresses can be linked to physical persons, seriously compromising their privacy.

Anonymizing the communication layer is thus a necessary measure to protect the privacy of users, and protect computer systems against traffic analysis. They also support anonymization techniques at the application layer, such as anonymous credentials, elections and anonymous cash.

Contributions

Anonymous communications research has matured to the point that new systems must imperatively take into account the existing literature and ensure that they are not weak under known attacks and models. The aims of this survey has been to present a road map of the most important systems-concepts, and the key refinements they have been subject to.

As in any mature field new designs will inevitably have to mix and match from elements already present, in older systems, to best match their environment. Designs tailored to peer-to-peer systems or telephony are a prime example of this. Those systems are also a prime example of the care that researcher must exert then mixing and matching ideas: as we show in this survey, anonymous communications are fragile, and even simple modifications may lead to traffic analysis attacks.

4.3 Accountable anonymous communication infrastructure

- Claudia Diaz and Bart Preneel. *Accountable Anonymous Communication.* To appear as Chapter of Security, Privacy and Trust in Modern Data Management, 15 pages, Springer.

Summary

We motivate in [DP06] the need for anonymity at the communication layer and describe the potential risks of having traceable communications. We then introduce the legal requirements on data retention and motivate the need for revocability of anonymity upon the request of law enforcement.

We describe the main building blocks for anonymous communication and for anonymity revocation. We explain how these building blocks can be combined in order to build a revocable anonymous communication infrastructure that fulfills both privacy and law enforcement requirements.

Background

Privacy is increasingly understood as an interdisciplinary subject. Legal, political and social considerations must be taken into account in the design of viable technical solutions that can be implemented at a large scale and accepted by the various players: citizens, governments, companies, etc. Anonymity and identity management technologies are powerful tools to protect privacy. Nevertheless, their potential for abuse is a factor that hinders the development and implementation of privacy enhancing systems at a large scale.

Contributions

This paper discusses the requirements for a large scale anonymous infrastructure, and proposes a model to comply with them. Our model is built by combining existing technologies and distributes trust among various entities. Anonymity infrastructures that protect the privacy of millions of individuals can only be possible if extreme care is taken in balancing the requirements of a multiplicity of interacting entities with sometimes conflicting interests. Otherwise, anonymity systems face the threat of remaining marginal in an environment in which privacy violations become ever more common.

4.4 Two sided statistical disclosure attack

- George Danezis, Claudia Diaz and Carmela Troncoso. *Two-Sided Statistical Disclosure Attack.* Accepted paper at Privacy Enhancing Technologies 2007 (PET'07), 15 pages.

Summary

We introduce in [DDT07] a new traffic analysis attack: the Two-Sided Statistical Disclosure Attack, that tries to uncover the receivers of messages sent through an anonymizing network supporting anonymous replies. We provide an abstract model of an anonymity system with users that reply to messages. Based on this model, we propose a linear approximation describing the likely receivers of sent messages. Using simulations, we evaluate the new attack given different traffic characteristics and we show that it is superior to previous attacks when replies are routed in the system.

Background

The field of anonymous communications started in 1981 with David Chaum's mix [Cha81]. A mix is a relaying router that ensures, through cryptography and reordering techniques, that input messages cannot be linked to output messages, therefore providing anonymity. Yet, it has been known for some time that anonymity systems, not offering full unobservability, are insecure against Long Term Disclosure [AK03] and Statistical Disclosure Attacks [Dan03].

Despite the level of protection that mix networks provide, they still leak some information. An external observer is able to find out the identities (or at least network addresses) of mix users sending or receiving messages, as well as the exact time messages are sent and received. We can find in the literature a powerful family of Disclosure Attacks [AK03], first proposed by Kesdogan et al. [KAP02]. These attacks allow an observer to learn the correspondents of each user and, in the long run, de-anonymize their messages. A different style of attack, the Statistical Disclosure Attack (SDA) [DS04], considers the same user behavior and communication model, but reduces the computation complexity by using statistical models and approximations to reveal the same information to an attacker.

Contributions

In this work, we extend Statistical Disclosure Attacks [Dan03] (SDA) in order to model user's behavior that deviates from the standard model considered so far in the literature. We consider that users not only *send messages* to a list of contacts, but also *reply to received messages* with some probability. Despite the real-world significance of modeling systems that allow anonymous replies, this is the first in-depth study of their security.

The Two-sided Statistical Disclosure Attack (TS-SDA) is the first traffic analysis attack to be explicitly targeted at anonymous communication systems that allow anonymous, and indistinguishable, replies. It takes into account the existence of replies and the timing of messages to estimate the correspondents of a target user and to trace the messages they send.

An adversary deploying our *Two-sided Statistical Disclosure Attack* (TS-SDA) takes into account the fact that some messages sent by a target user Alice are replies, in order to infer information on the set of Alice's contacts, and trace individual messages more effectively. This is done by combining information from sender and receiver anonymity sets when tracing replies.

We show through simulations that the Two-sided Statistical Disclosure Attacks give much better results than the traditional Statistical Disclosure Attacks, when tracing anonymized traffic that contains replies. We also evaluate how the effectiveness of our attacks is influenced by users' behavior (e.g., how often users reply, or how long it takes them to reply).

4.5 Future work

Many issues concerning anonymous communications remain unresolved. In ADAPID, we will continue to investigate techniques for building anonymous communication infrastructures and for testing the robustness of existing designs towards attacks.

Two of the topics we intend to investigate in the near future are peer-to-peer anonymous communication networks for low-latency voice communications (voice over IP), and user models for anonymous email usage.

Chapter 5

Secure and Privacy-Enhanced Storage

5.1 Introduction to storage

Nowadays, documents are created as (or transformed into) digital records. This brings the necessity of having digital repositories where this information can be stored and accessed in a secure way. Traditionally, the primary goal of storage security is to ensure the *CIA* properties:

- **Confidentiality**: no one should have access to data unless specifically authorized.
- **Integrity**: consists in keeping data consistency against accidental/malicious attacks. This implies that no data has been subjected to unauthorized modification. Ensuring integrity has two aspects: data modification prevention and data modification detection (which includes recovery of the original data).
- **Availability**: a user must be able to access information within an acceptable time period, without monopolizing the storage source and in an understandable representation.

The maintenance of these properties should be done without losing any *performance* in the system. The security measures applied to achieve these three objectives should not become an excessive overhead, i.e. the system should – within reasonable limits – be as usable as if no security mechanism was implemented.

Storage is one of the essential building blocks for most of the applications in e-government or e-health. Digital repositories are needed to preserve medical records, government archives, etc. The features mentioned above are crucial for the correct operation of any archival system, but they are not the only ones that need to be taken into account when building a secure archiving system. As more personal data are kept, privacy protection of the data subjects becomes a priority, even more if we take into account the amount of personal information recorded on Belgian e-ID cards. Also from a legal point of view, privacy protection is of major importance, the European Data Protection directive (1995/46/*EC*) addresses how processing of data should be done and how data should be protected and released. Mechanisms to insert or retrieve data from a repository in a privacy friendly way are needed, and additionally special confidentiality protection is necessary when sensitive information is stored.

Moreover, as e-ID cards are used to electronically sign digital documents, storage methods should be available ensuring that these documents are a functional equivalent for traditional paper documents with a manual signature. For this to become real, we need to be able to:

- prove the validity of digital signatures well into the future when, most probably, the certificates related to it are no longer available (or no longer valid), and
- prove that the content of the signed document has not changed during storage.

Another desirable property for an archival system would be the ability to keep the confidentiality of a record over time, as the algorithms used to hide its contents from unauthorized users will most probably be broken in the future. Finally, archival services should provide a way to deal with the obsolescence of software, allowing the format of the records they store to be transformed or providing mechanisms to read them when released at a later date.

In the next section we put forward the applications of our basic research result in the ADAPID scope. Sections 5.2 to 5.4 present the basic work done concerning secure storage with mainly two lines of research:

- *Steganographic file systems*: file systems which allow the user to plausibly deny the existence of files in a repository, see Sect. 5.2. We have developed an attack on the most recent (to the best of our knowledge) proposition of a Steganographic file system (see Sect. 5.2), and designed a new approach which is described in Sect 5.3.
- *Private search*: allows for searching in streaming data with encrypted keywords, and retrieving the search results without leaking any information to the searching party. This work is explained in Sect. 5.4.

At the end of the chapter, in Sect. 5.5 we outline work in progress and future work which we plan to do in this area.

Relevance to ADAPID

The first two papers we present in this deliverable (Sect 5.2 and Sect. 5.3) show our work in Steganographic file systems, see Sect. 5.2. Steganographic file systems allow the user to hide groups of files (levels) with different degrees, e.g. some files may be accessible only to him personally, while others may be visible also to co-workers, superiors or relatives. This has several applications in the real world when we talk about personal safety. It can be used by NGOs, journalists, politicians,. . . who have sensitive information (e.g, human rights, war investigations, etc.) in order to protect themselves against local authorities or other attackers. These type of systems also have applications in e-government and e-health, when even the number of confidential records that a repository has could be sensible information. Thus, to avoid any leak of information, separation between low and high level users needs to go further than simple encryption, which can be provided by Steganographic file systems.

In the first paper (see Sect 5.2) we have successfully deployed traffic analysis attacks on the latest proposal of a Continuously-Observable Steganographic file system. This demonstrates that the current state of the art is far away from a definite solution. While developing these attacks we have gained considerable experience that has helped us to build a better approach, which we describe in Sect. 5.3.

The third paper (Sect. 5.4) presents our results in the field of private search. They can be used in a wide variety of applications. For example, these techniques can be used in the context of e-health by patients accessing a database with information on diseases, so that the database owners cannot infer which diseases may be suffered by the person who is making the query. In the case of e-government, these techniques can also be used to provide privacy-enhanced access to public databases; e.g., citizens looking for the criteria to be eligible for social services would not need to reveal the specific conditions they are subject to – for instance, disability.

5.2 Traffic Analysis Attacks on a Continuously-Observable Steganographic File System

- Carmela Troncoso, Claudia Diaz and Bart Preneel. *Traffic Analysis Attacks on a Continuously-Observable Steganographic File System.* This paper was accepted for publication in April 2007 at the 9th Information Hiding Workshop.

Summary

In this paper [TDP07] we analyze a previous proposal on a Continuously-Observable Steganographic file system, StegFS [PTZ03, ZPT04], in which the authors claim that the system resists attackers who can continuously monitor accesses to the files storage. This proposal relies on dummy updates and relocations of data that are supposed to conceal accesses to the hidden files. We have demonstrated that this system is vulnerable to traffic analysis attacks.

Background

The goal of a steganographic file system is to protect the user from compulsion attacks in which the user is forced to hand over file decryption keys under the threat of legal sanctions or physical intimidation. In order to achieve this goal, the steganographic file system must conceal the files it stores, so that the user can plausibly deny their very existence. A Continuously-Observable steganographic file system, aims to conceal the existence of files, even for an attacker that is able to permanently monitor the storage unit and accesses operations.

Zhou et al. proposed in [ZPT04] mechanisms to hide data accessing in their StegFS [PTZ03] against this attack model. In the system model (Fig. 5.2) of [ZPT04], users send the file requests to a trusted agent over a secure channel. The agent translates these requests into I/O operations on the repository, and returns the results to the user. Whenever there is no user activity, the agent performs dummy I/O operations.

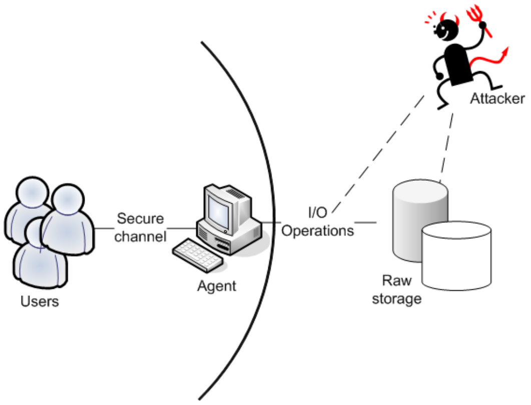


Figure 5.1: System model

Contributions

We have developed traffic analysis attacks which are effective against the file update mechanism. Our attacks succeed in revealing the existence and location of all the hidden files, depriving the user of plausible deniability. We have described the theory behind the attacks, and discussed the impact of the system's parameters on their effectiveness.

Our results show that traffic analysis attacks against [ZPT04] successfully distinguish file and dummy updates, and find the locations of hidden files. The two key weaknesses in the update algorithm proposed in [ZPT04] are:

- Blocks are rarely relocated, and when they are, their new location appears next to the old one in the history of accessed locations. This greatly reduces the uncertainty on the possible locations to which block contents may have been moved.
- While the dummy updates select block locations uniformly at random, multi-block file updates generate correlations between accessed locations that could not have been plausibly generated at random.

The traffic analysis strategies presented in this paper show that introducing "a bit of randomness" is not sufficient to effectively conceal user access to files in a steganographic file system. More sophisticated mechanisms are required in order to design a traffic analysis resistant steganographic file system.

We have also programmed a Python simulation of the attack, and obtained empirical results that confirm the vulnerability of [ZPT04] against traffic analysis. Our (non-optimized) implementation successfully finds most of the files hidden in the repository, and more efficient implementations could further increase the accuracy of the attacks.

5.3 Observable Steganographic File Systems using Pool Mixes and Universal Re-encryption

- Claudia Diaz, Peter Fairbrother and Carmela Troncoso. *Observable Steganographic File Systems using Pool Mixes and Universal Re-encryption.* 18p. This paper was accepted for publication in April 2007 at the 7th Privacy Enhancing Technologies Workshop.

Summary

In this paper [DFT07] we present a new way of building a steganographic file system (see Sect. 5.2), where files are kept in a repository which can be continuously monitored by an adversary. We present an overall design and a set of techniques which can be used to build a steganographic file system resistant to continuous observation.

For this purpose we combine universal re-encryption, pool mixes, erasure codes, and cover traffic. We demonstrate the semantic security of the universal re-encryption construction used, and introduce a mechanism to counter rollback attacks where the adversary recreates the file system at any previous state.

Background

Our system comprises the user Bob, Bob's computer, and a remote block storage device, see Fig. 5.2. In addition to the normal user software, Bob's computer contains a pool of encrypted storage blocks, a block location table, encryption/ decryption and re-encryption software, and a block selector (not shown). In our system user files are hidden in encrypted form amongst blocks of random data, as files encrypted in a semantically secure cipher are indistinguishable from random data.

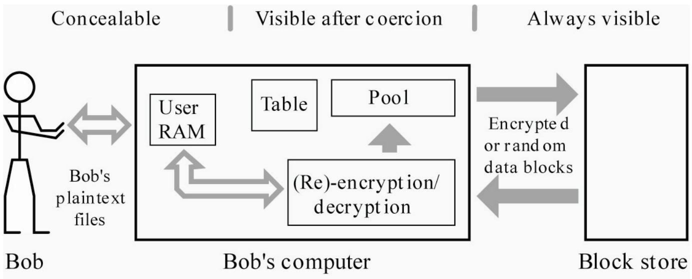


Figure 5.2: System model

In this paper we have used a combination of the following techniques:

- *Universal re-encryption* [Fai04, PGS04], allows re-encryption of ciphertexts without knowledge of the original encryption key, and is used here to change the appearance of encrypted data without actually modifying their plaintext contents while giving unlinkability between old and new versions of blocks (a semantically securely re-encrypted ciphertext is indistinguishable from the old ciphertext or a new ciphertext).

- *Erasure codes* [Riz97] convert an original m-block file in $n > m$ encoded blocks such that any subset m of these blocks suffices to recover the file. Its use in our system is twofold: first it improves the persistence of files in the system (high security file blocks appear as empty when the user is logged in with a low security level and user data may be lost when these blocks are overwritten) and second, they help complicating correlations between read operations on a file.

- *Pool mixes* are a well-known technique used to anonymize email traffic [Cot, DDM03]. The seminal idea of mixes was proposed by Chaum in [Cha81]. In this paper, we present a novel application for the pool mix: it is used to hide data access in an observable steganographic file system, making it resistant to traffic analysis attacks. At regular intervals a block chosen by the selector is read from the storage device, then the location in the storage space which contained that block is overwritten with a block chosen at random from the pool.

- *Dummy traffic* refers in this context to automatically generated access to the repository. In order to provide unobservability, we present a strategy that makes the pattern of access to raw storage locations in idle periods (dummy traffic) indistinguishable from the pattern generated by user requests.

Contributions

The contributions of this paper are the following: we present a new design for an observable steganographic filing system; demonstrate the semantic security of the universal re-encryption construction used, proving resistance to passive attacks and introducing methods to resist active chosen plaintext and ciphertext attacks; introduce nonces in order to prevent rollback attacks; present a new application for mixes, justify the selection of the chosen pool mix, and show that it provides optimal unlinkability for a given available amount of local storage (pool size); explain the differences between traffic analysis for mixes in the context of anonymous communication and in file systems; develop new traffic analysis techniques that take into account the feedback introduced by the limited number of possible locations of a block, which can be applied to other designs utilizing another kind of mix (e.g., Stop-and-Go mixes [KEB98]) and different dummy traffic strategies. Finally, we discuss the limits of unobservability in this system, and explain the impact of the system's parameters on these limits.

5.4 Private Search

Our research results have been published in two papers. The first, presented at the Dagstuhl Seminar on Anonymous Communication, proposed a preliminary version of our solutions, that were later improved in the publication at the International Conference on Financial Cryptography and Data Security.

- George Danezis and Claudia Diaz. *Space-Efficient Private Search.* To appear in the Proceedings of the International Conference on Financial

Cryptography and Data Security (FC'07), 15 pages, LNCS (in print), Springer, 2007.

- George Danezis and Claudia Diaz. *Improving the Decoding Efficiency of Private Search.*In Anonymous Communication and its Applications, number 05411 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fur Informatik (IBFI), Schloss Dagstuhl, 11 pages, 2006.

Summary

Private keyword search is a technique that allows for searching and retrieving documents matching certain keywords without revealing the search criteria. We improve the space efficiency of the Ostrovsky *et al.* Private Search [OS05] scheme, by describing methods that require considerably shorter buffers for returning the results of the search [DD06, DD07a]. Our basic decoding scheme *recursive extraction*, requires buffers of length less than twice the number of returned results and is still simple and highly efficient. Our extended decoding schemes rely on solving systems of simultaneous equations, and in special cases can uncover documents in buffers that are close to 95% full. Finally we note the similarity between our decoding techniques and the ones used to decode rateless codes, and show how such codes can be extracted from encrypted documents.

Background

Private search allows for keyword searching on a stream of documents (typical of online environments) without revealing the search criteria. Its applications include intelligence gathering, medical privacy, private information retrieval and financial applications. Financial applications that can benefit from this technique are, for example, corporate searches on a patents database, searches for financial transactions meeting specific but private criteria and periodic updates of filtered financial news or stock values.

Rafail Ostrovsky *et al.* presented in [OS05] a scheme that allows a server to filter a stream of documents, based on matching keywords, and only return the relevant documents without gaining any information about the query string. This allows searching to be outsourced, and only relevant results to be returned, economising on communications costs. The authors of [OS05] show that the communication cost is linear in the number of results expected. We extend their scheme to improve the space-efficiency of the returned results considerably by using more efficient coding and decoding techniques.

Contributions

Our first key contribution is a method called *recursive extraction* for efficiently decoding encrypted buffers resulting from the Ostrovsky *et al.* scheme. The second method, based on solving systems of linear equations, is applied after recursive extraction and allows for the recovery of extra matching documents from the encrypted buffers. Recursive extraction results in the full decoding of buffers of length twice the size of the expected number of matches, and has a linear time-complexity. Shorter buffers can also be decrypted with high probability. Solving the remaining equations at colliding buffer positions allows for even more documents to be retrieved from short buffers, and in the special case of documents only matching one keyword, we can decode buffers that are only 10% longer than the expected matches, with high probability. We present simulations to assess the decoding performance of our techniques, and estimate optimal parameters for our schemes.

In this work we also present some observations that may be of general interest beyond the context of private search. We show how arrays of small integers can be represented in a space efficient manner using Pailler ciphertexts, while maintaining the homomorphic properties of the scheme. These techniques can be used to make private search more space-efficient, but also implement other data structures like Bloom filters, or vectors in a compact way. Finally we show how rateless codes, block based erasure resistant multi-source codes, can be extracted from encrypted documents, while maintaining all their desirable properties.

5.5 Work in progress and Future Research

We are currently working on the design of a Long Term Archive Service that allows signatures to be verified in the future, and that is able to prove that the contents of files stored using this service have not been modified. Furthermore, we will develop mechanisms to provide Long-Term Confidentiality and to allow transformations on the format of a record without losing the integrity or the validity of the signature.

We will also study the state of the art of storage in the literature, both in the traditional way (confidentiality, integrity, availability) and looking for advanced properties (enhancing privacy, long term properties). This will result in a survey of the existing archiving systems.

A third line of research in storage applications is the study of negative databases [EAH$^+$06]. A negative database is a representation of a set of records (the positive database) that allows its holder to test whether particular entries are present in it, but makes it very hard to efficiently enumerate all entries. We are working in a more efficient implementation of these databases than the one proposed originally in [EAH$^+$06], based on cryptographic hash functions and the family of Diffie-Hellman assumptions.

Chapter 6

Techniques for Distribution of Trust

6.1 Introduction

In many scenarios it is desirable to distribute trust among a group of entities, so that certain specified groups of them are able to perform an operation, but smaller, malicious subsets of them cannot do any harm. These questions are addressed by concepts like secret sharing schemes (SSS), verifiable secret sharing (VSS) schemes, and multi-party computation (MPC). The concept of secret sharing has been introduced independently both by Shamir and Blakley as a tool to protect a secret information from being lost or exposed. A so-called dealer shares the secret among the set of members, called players or participants, in such a way that only certain specified subsets of players (access structure) are able to reconstruct the secret, while smaller subsets have no information about this secret at all (in a strict information theoretic sense). A stronger version called verifiable secret sharing (VSS) has been introduced in order to achieve better robustness against malfunctions or misuse.

In secure Multiparty Computation (MPC): we have n players, at most t of which are bad (i.e., do not follow the protocol, may leak information); there are inputs $x_1, ..., x_n$, distributed among the players, and there is a function $f(x_1, ..., x_n)$ that the n players wish to compute. The goal is to construct a protocol to evaluate f which is correct and t-secure, i.e., no set of t bad players can prevent good players from computing the correct output nor can they learn any information that they are not supposed to know. In addition, the protocol must of course be as efficient as possible. The interest in MPC comes from the fact that all security related tasks, many of which are today handled using paper work, can in principle be handled electronically using a general MPC protocol.

Threshold secret sharing, where all participants play an equal role, has found numerous applications such as key escrow, distributed file storage and distributed computation. However, threshold secret sharing does not allow to distinguish participants based on trust or authority, whereas in many real-life scenarios, hierarchies are required. Consider an example from the military. Let the secret be the "nuclear button" of a country and suppose that it can only be accessed by two ministers, or a minister and a general, but not by two generals. Then a threshold scheme is clearly not suitable, since any two generals could pool their shares together to pass the access control. In one of the papers below we consider exactly this scenario, namely hierarchical threshold secret sharing.

Relevance to ADAPID

As we have explained above, in many scenarios it is wise to divide the secret information among several parties instead of trusting just a single server - there exists always the risk that it becomes corrupted or crashed. This also applies to applications which involve e-ID cards. In this respect, research on the main building blocks for distribution of trust (like secret sharing schemes and multi-party computation) is important.

6.2 On a Relation between Verifiable Secret Sharing and a Class of Error-Correcting Codes

- Ventzislav Nikov, Svetla Nikova, *On a Relation between Verifiable Secret Sharing and a Class of Error-Correcting Codes*, In Proceedings of the International Workshop on Coding and Cryptography (WCC 2005), pp. 372-382, 2005, Springer-Verlag, LNCS 3969, 2005, pp. 275-290.

Summary

In this paper [NN05a] we try to shed a new insight on Verifiable Secret Sharing Schemes (VSS). We define a new "metric" and using it we define a very particular class of codes that we call *error-set correcting codes*, based on a set of forbidden distances which is a monotone decreasing set. We establish a link between Secret Sharing Schemes (SSS) and the error-set correcting codes and we give a necessary and sufficient condition for the existence of linear SSS (LSSS).

Background

The concept of *secret sharing* was introduced by Shamir [Sha79] as a tool to protect a secret from getting exposed or from getting lost. It allows a so-called *dealer* to share a secret among the members of a set $\mathcal{P}$, which are usually called *players* or *participants*, in such a way that only certain specified subsets of players are able to reconstruct the secret (if needed) while smaller subsets have no information about this secret at all (in a strict information theoretic sense).

We call the groups who are allowed to reconstruct the secret *qualified* and the groups who should not be able to obtain any information about the secret *forbidden*. The set of qualified groups is denoted by Γ and the set of forbidden groups by Δ. The tuple (Γ, Δ) is called an *access structure* if $\Gamma \cap \Delta = \emptyset$. If the union of Γ and Δ is equal to $2^{\mathcal{P}}$ (so Γ is equal to Δ^c, the complement of Δ), then we say that access structure (Γ, Δ) is *complete* and we denote it just by Γ. In the sequel we shall only consider complete, monotone access structures.

It is common to model cheating by considering an *adversary* $\mathcal{A}$ who may corrupt some of the players. The adversary is characterized by particular subset Δ_A of Δ, called *adversary and privacy structures* [MH00] respectively, which are monotone decreasing structures. The players which belong to Δ are called also *curious* and the players which belong to Δ_A are called *corrupt* or *bad*.

Contributions

We first define a new "metric" (with slightly different properties than the standard Hamming metric). Using this metric we define a very particular class of codes that we call *error-set correcting codes*, based on a set of forbidden distances which is a monotone decreasing set. Next we redefine the packing problem for the new settings and generalize the notion of error-correcting capability of the error-set correcting codes accordingly (taking into account the new metric and the new packing). Then we consider burst-error interleaving codes proposing an efficient burst-error correcting technique, which is in fact the well known VSS and Distributed Commitments (DC) pair-wise checking protocol and we prove the error-correcting capability of the error-set correcting interleaving codes.

Using the known relationship, due to Van Dijk, between a Monotone Span Program (MSP) and a generator matrix of the code generated by the suitable set of vectors, we prove that the error-set correcting codes in fact has the allowed (opposite to forbidden) distances of the dual access structure of the access structure that the MSP computes. We give an efficient construction for them based on this relation and as a consequence we establish a link between Secret Sharing Schemes (SSS) and the error-set correcting codes.

Further we give a necessary and sufficient condition for the existence of linear SSS (LSSS), to be secure against (Δ, Δ_A)-adversary expressed in terms of an error-set correcting code. Finally, we present necessary and sufficient conditions for the existence of a VSS scheme, based on an error-set correcting code, secure against (Δ, Δ_A)-adversary.

Our approach is general and covers all known linear VSS/DC. It allows us to establish the minimal conditions for security of VSSs. Our main theorem states that the security of a scheme is equivalent to a pure geometrical (coding) condition on the linear mappings describing the scheme. Hence the security of all known schemes, e.g. all known bounds for existence of unconditionally secure VSS/DC including the recent result of Fehr and Maurer, can be expressed as certain (geometrical) coding conditions.

6.3 On Secret Sharing Schemes

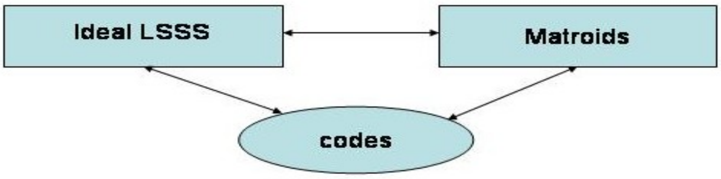
- Svetla Nikova, Ventzislav Nikov, In Proceedings of the Contact Forum *Coding theory and cryptography*, Royal Flemish Academy for Science and the Arts, Brussels, October 7, 2005, pp. 59-75.

Summary

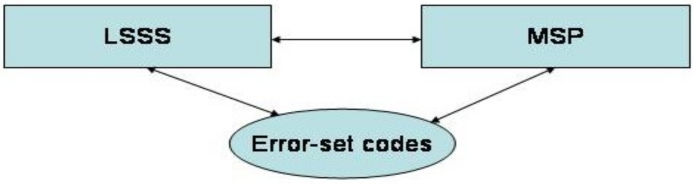
In this paper [NN05b] we present an overview of some approaches for building secrets sharing schemes based on well studied objects like matroids and error-correcting codes on one hand. We start with introducing the Shamir's polynomial scheme. Then we talk about the relations between ideal Secret Sharing, matroids and error-correcting codes. On the other hand we introduce linear SSS for general access structures and we explain the approach by Cramer, Damgard and Maurer based on Monotone Span Programs. We complete by considering error-set codes as a generalization of the notion of codes.

Contributions

In this paper we give an overview of some of the approaches for building secret sharing schemes. First we present two approaches based on well studied objects, namely, error correcting codes - to build threshold SSS and matroids - to build ideal linear SSS. These approaches are efficient but not general. The first known techniques for building general SSS lead to exponential number of shares. Moreover these approaches are not suitable for building more complex protocols like verifiable SSS or multi-party computation on the top of SSS.



In 1993 MSP have been introduced as an algebraic model of computation. Later it has been proven that for any general access structure can be build linear SSS based on MSPs. Moreover the number of shares in the worst case is super-polynomial. On the other hand MSPs can be considered as a generalization of matroids. Recently we have introduced error-set codes as a generalization of the notion of codes. By means of error-set codes and MSPs one can build general SSS. Moreover this approach is general and can easily be applied to more complex protocols, which use SSS as a building block.



Finally we have shown that the relations that are valid between ideal linear SSS, matroids and error-correcting codes are also valid between their generalizations - LSSS, MSP and error-set codes, respectively (see the figures).

6.4 Strongly Multiplicative Hierarchical Threshold Secret Sharing

- Emilia Käsper, Ventzislav Nikov, and Svetla Nikova, *Strongly Multiplicative Hierarchical Threshold Secret Sharing*, accepted paper at the International Conference on Information Theoretic Security 2007 (ICITS'07).

Summary

We consider multi-party computation (MPC) in a hierarchical setting, where participants have different capabilities depending on their position in the hierarchy [KNN07]. First, we give necessary conditions for multiplication of secrets in a hierarchical threshold linear secret sharing scheme (LSSS). Starting with known ideal constructions, we then propose a modified scheme with improved multiplication properties. We give sufficient conditions for the new scheme to be (strongly) multiplicative and show that our construction is almost optimal in the number of required participants. Thus, we obtain a new class of strongly multiplicative LSSS with explicit ideal constructions. Such LSSS are also useful outside the MPC setting, since they have an efficient algorithm for reconstructing secrets in the presence of errors.

Background

In threshold secret sharing, a secret is shared amongst n participants and can only be reconstructed by more than t of them together. Such schemes have found numerous applications such as key escrow, distributed file storage and distributed computation. However, in threshold schemes, all participants play an equal role and cannot be distinguished according to trust or authority, whereas in many real-life situations, hierarchies are required. Consider an example from the military. Let the secret be the "nuclear button" of a country and suppose that it can only be accessed by two ministers, or a minister and a general, but not by two generals. In this case, a 2-out-of-n threshold scheme is clearly not suitable, since any two generals could pool their shares together to bypass access control. Secret sharing schemes that take into account hierarchies were the first non-threshold schemes considered in the literature [Sha79, Sim90, Bri89].

In this paper, we investigate the *multiplicativity* of hierarchical schemes. Multiplicativity allows participants, holding shares of two secrets s and s', to privately compute shares of the product ss' without revealing the original secrets. *Strong* multiplicativity further guarantees that in the presence of an adversary, honest participants can still compute such shares. A simple solution for multiplication of secrets exists for the Shamir threshold scheme [GRR98]. In general, however, it is not known how to efficiently construct a strongly multiplicative linear secret sharing scheme (LSSS) with the desired non-threshold access structure. Thus, we tackle a more specific problem and prove strong multiplicativity for a class of access structures.

Strongly multiplicative schemes turn out to be useful even outside the context of multiplying secrets, since they are resistant to errors in shares. Although in any LSSS with a special access structure (called Q^3-access structure), the secret is uniquely determined even if the shares submitted by corrupted participants contain errors, it is not known how to locate such errors efficiently. An efficient secret reconstruction algorithm is only known for strongly multiplicative LSSS [CDG$^+$05]. This implicit "built-in" verifiability makes strongly multiplicative schemes an attractive building block for multi-party computation (MPC) protocols.

Contributions

Strongly multiplicative secret sharing schemes are used in multi-party computation to obtain error-free multiplication unconditionally secure against an active adversary. However, enforcing the multiplication property is in general expensive and few efficient non-threshold examples are known.

In this paper, we analyze the multiplicativity of two important families of hierarchical secret sharing schemes: threshold schemes based on conjunction and disjunction of conditions. First, we prove necessary conditions for multiplicativity. Then, we look at constructions based on Shamir threshold secret sharing and show that they are always (strongly) multiplicative whenever these necessary conditions are fulfilled. The constructions are not ideal but have a reasonable information rate when the number of levels in the hierarchy is small. Next, we investigate ideal constructions and propose a new conjunctive scheme based on the Tassa scheme [Tas04]. We prove sufficient conditions for (strong) multiplicativity of the modified scheme. The conditions are not tight but we demonstrate that the gap is quite small, i.e, the construction is close to optimal. Finally, we note that our modified scheme actually has better multiplicative properties than the original scheme of Tassa.

Thus, as a result of our analysis and our improvements to existing designs, we describe a big class of strongly multiplicative secret sharing schemes that have an efficient ideal construction.

Chapter 7

Policies and Digital Rights Management

7.1 Introduction

Policies are important in privacy friendly applications. They define the rules with which the applications will abide. Policies serve different purposes:

- They **convey information** about an application. For instance, a service provider informs the user (client) through its privacy policy what information is gathered, for what purpose, how and for how long it will be retained, etc. This does not mean that the service provider will abide by these rules; however, the service provider may be held liable for breaching the rules specified in the policy.

- Policies make applications **more flexible**, since the rules are not hard-coded in the applications themselves. This means that users can more easily customize an application to their needs. The customization can take context information into account; e.g. the identity of the service provider, the time of the day, the location of the client (user), the hardware used to run the application, etc.

  The owner of a resource can protect its usage by specifying the authorized operations in a policy. For instance, the owner of personal data can specify to whom, in what form and under which circumstances that data can be disclosed. On the downside, the use of policies may make it difficult to understand what exactly is allowed and what is not. Therefore, powerful (graphical) tools should help the user to set up and interpret his policies.

  These policies need to be enforced; hence, appropriate enforcement mechanisms must be available to the application designer. Before every action that needs authorization, a policy enforcement entity must intervene and request authorization from one or more policy decision entities (these are entities that interpret the policies and decide whether or not the action is allowed).

- Policies allow for **automation**; hence, they reduce the need for user interaction. Too many interactions with the user inevitably leads to a reduction of the security and privacy. Policies allow applications to authorize or disallow certain actions without user interaction. Only when the policy explicitly states that user's approval is required, or in circumstances that fall outside the policy, user interaction will occur. This way, users will stay more alert for warnings from the system.

In the context of the ADAPID project, we will need three different classes of policies. They serve different purposes (as was explained in the previous paragraph).

- **Privacy policies** are used by service providers, to describe the information that users will have to provide in order to get access to a particular service. The policies also define which authorities should have certified that information, and in which form the information should be provided (e.g. cleartext, verifiably encrypted, ...). Finally, these policies announce how the requested data will be processed, where and for how long that data will be stored and to which other parties the information can be disclosed, as this is required by the law safeguarding the individual privacy.

  Based on the rules specified in the policy, the client can decide whether or not these rules are strict enough to continue the interaction with the service provider. Also, the policy can be kept by the user and used in a court case if a breach of this policy is detected.

- Through **disclosure policies** or **privacy preferences**, the user can restrict the information (or properties thereof) that will be divulged to service providers. For instance, the policy may state that the user's yearly income may only be disclosed to financial institutions and the IRS.

  These policies are necessary to customize the application. They restrict the privacy policies that are acceptable. Clearly, a match must be found between the privacy policy and the disclosure policy. Moreover, disclosure policies also help to automate the selection and usage of credentials without too much user interaction.

- **Access policies** restrict the access to and the usage of resources. They specify what operations are authorized under what circumstances.

Sometimes, they are inextricably bound up with the resources they protect, in which case they are called **sticky policies**.

Access policies may be used by service providers to restrict access to personal data received from clients and, hence, to ensure that the privacy policy is not violated. Also, Digital Rights Management systems (DRM) may use access policies as part of the license or include such a policy in the protected resource.

Today, several privacy, disclosure and access policy languages exist. However, they are not well suited to deal with anonymous credentials (see also 'Background information' of section 7.2). Hence, the first step taken in this research track, is the extension of these policy languages.

DRM or **Digital Rights Management** is the generic term for techniques used by publishers or copyright owners to control access to and usage of digital data or hardware. It may also refer to the restrictions associated with a specific instance of a digital work or device.

Much like a copyright owner wishes to control access to and use of his works, data subjects would like to control access to and use of data that refers to them, especially with regard to sensitive personal information. In this context, DRM techniques could be used to achieve such control.

In this research track, we explore when and how DRM techniques can be introduced to protect personal data.

Relevance to ADAPID

Users will have hundreds of credentials, certified by many different authorities. Legal obligations and data assurance requirements compel service providers to formulate a privacy policy. To avoid unnecessary interaction with the user, a disclosure policy may guide the system to continue or abort the interaction with a service provider and to select the appropriate credentials and prove, show or hide embedded attributes. As a first step, existing policy languages need to be extended to support credentials and other types of certified data.

DRM techniques can provide an additional means to protect access and usage of the user's personal data. Since privacy protection is the main theme in ADAPID, this research is relevant and innovative. Two examples are given:

- In the E-health application domain, EHRs (Electronic Health Records) can be protected using DRM, such that the patient can determine in a flexible way the usage rights for different doctors; e.g. a doctor can get the right to access the EHRs until the end of the year. Delegation of rights can be described as well such that the doctor can delegate the access rights (partly) to other accredited persons such as nurses.

- DRM can be used to enforce sticky policies associated with credentials. Sticky policies are specified by the credential issuer, but without DRM, it is hard to ensure that they are enforced properly at the client side. A sticky policy could for instance state that the credential may only be shown starting from some date, and only to some predetermined set of verifiers.

7.2 Privacy friendly information disclosure

- Steven Gevers and Bart De Decker, Privacy friendly information disclosure, On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops (Meersman, R. and Tari, Z. and Herrero, P., eds.), vol 4277, Lecture Notes in Computer Science, pp.636-646, 2006.
- Steven Gevers, and Bart De Decker, Automating privacy friendly information disclosure, K.U.Leuven, Department of Computer Science, Report CW 441, April, 2006.

Summary

When using electronic services, people are often asked to provide personal information. This raises many privacy issues. To gain the trust of the user, service providers can use privacy policy languages such as P3P to declare the purpose and usage of this personal information. User agents can compare these policies to privacy preferences of a user and warn the user if his privacy is threatened.

This paper extends two languages: P3P and APPEL. It makes it possible to refer to certified data and credentials. This allows service providers to define the minimal level of assurance. It is also shown how different ways of disclosure (exact, blurred, verifiably encrypted, ...) can be specified to achieve more privacy friendly policies.

Next, the paper describes a privacy agent. This agent makes use of the policies to automate privacy friendly information disclosure.

Background

Service providers can use privacy policy languages to specify the purpose and usage of personal information. User agents can compare these policies to the user's privacy preferences and warn the user if his privacy is threatened. Two well known privacy languages are P3P (The Platform for Privacy Preferences[Pla]) and APPEL (A P3P Preference Exchange Language[A P]). The former is used for privacy policies, the latter for privacy preferences.

In the current version of these languages one can describe different types of information; however, it is not possible to specify by which authority this information has to be endorsed or certified. Hence, it is possible to specify that the service provider will request name and address of the the user, for what purpose this information is needed, to which other parties name and address may be disclosed, and how long this information will be retained, but the policy cannot specify that this information should be sent signed by the national registry, or embedded in a valid certificate issued by a trusted party, or be disclosed during a show protocol of a credential in which these fields are embedded. Moreover, the languages cannot cope with proving properties of attributes (such as being over 18), or with information that is verifyably encrypted for a trusted third party.

Contributions

The paper extends P3P and APPEL with the different ways of disclosure and different types of information structures. This way, service providers can request a certain level of assurance. Also, the user's privacy can be better protected. The paper also describes how the policies can be used by a *privacy agent* to automate information disclosure.

7.3 A Flexible and Open DRM Framework

- Kristof Verslype and Bart De Decker, A Flexible and Open DRM Framework, Communications and Multimedia Security 2006, (Leitold, H. and Markatos, E., eds.), vol 4237, Lecture Notes in Computer Science, 2006, pp. 173-184.

Summary

Current DRM implementations rely on obfuscating the inner working of the DRM client. Moreover, the rights to consume content are rather device bound than person bound. The paper presents a first step towards an open DRM framework which is based on the security of its building blocks. The presented framework binds the right to consume content to persons instead of to devices. An extension of the current TPM specification is proposed to enhance the security of DRM clients.

The basis of the person binding solution is to have two types of licenses: a content license and a root license. A content license gives a specific consumer the right to perform some actions on DRM protected content. A root license is device bound and enables the consumer to use all his content licenses on that device. A reasonable extension of the current TPM specification is proposed in order to obtain a DRM framework that is hard to break.

Background

Current DRM implementations are closed source and most details are hidden; currently, the security of the DRM technologies relies on the secrecy of algorithms in the DRM client. Due to this approach, even the DRM technologies that are considered as the most mature and most secure have been broken[Win05].

Contributions

This paper is a first step towards more flexible DRM.

- The ability to consume content is bound to the consumers themselves, where currently, this is bound to one or more devices.
- The presented framework is flexible in the sense that it is based on building blocks that can be replaced if they are no longer considered appropriate. The required building blocks are not all equally mature, but we can expect that this will improve in the near future.

7.4 Work in Progress

The ADAPID-framework will eventually support privacy and disclosure policies. A special kind of disclosure policy is a so called **sticky policy**, which is attached to a credential and restricts its usage; it should not be possible to remove or adapt the policy, neither to disregard it. Special tamperproof hardware or DRM techniques might be helpful to realize these sticky policies.

By applying DRM techniques to e-health records, the patient will have better control over the use of his e-health records.

Chapter 8

Model-driven Approaches to adapID

8.1 Introduction

Model-based approaches are increasingly used in all stages of complex systems design. In recent years, research has evolved in the area of dependable system design through model-based approaches. In [ZMVK05], the authors use a modelling and simulation-based process to ensure that a Pump Control System adheres to specified safety and reliability requirements. While various methodologies have been designed to tackle such dependability issues (as discussed in Sections 8.2 and 8.3) , other emerging constraints, such as security, privacy and anonymity, need to be addressed. Indeed, as the Internet evolves from being a basic communication mechanism to more of a tool to effectively operate businesses and provide services, more sensitive information is being digitized, and various e-applications require the access to, manipulation, and transfer of such data, thus making it susceptible to third-party attacks, or even user abuses. This opens up a lot of these digitized processes to outsiders. The correct operation of these applications is fundamental and security/control is more than ever a crucial concern. Such e-systems must, therefore, be: 1) secure, and designed in such a way that agents (users or programs) can only perform actions that have been allowed; 2) privacy-sensitive, and effectively protect personal data from unauthorized access or disclosure.

At present, the identity and on-line behaviour of individuals is routinely recorded; users often have little knowledge of or control over such surveillance. Many anonymous applications offer unconditional anonymity to their users. However, this can provoke abusive behavior. Dissatisfied users will drop out or liability issues may even force the system to suspend or cease its services. Therefore, controlling abuse is as important as protecting the anonymity of legitimate users. However, designing such applications is no sinecure. The paper presented in Section 8.4 presents a methodology for designing controlled anonymous environments. The methodology generates a conceptual model that compromises between privacy requirements and control requirements. The conceptual model allows to derive performance and trust properties and easily maps to control mechanisms.

Similarly, to tackle security requirements of an e-ID application, the paper described in Section 8.5 proposes a multi-formalism based process geared towards security constraint analysis and verification. For this purpose, initial requirements are taken into consideration in the design of the application. This design model is then translated into a verification model, represented in a language which offers verification capabilities, and adequate tools for model checking. In Section 8.6, the paper summarizes a new language developed to described a view of the applications which is between requirements and verification model.

Finally, the thesis and paper introduced in Section 8.7 describes a model-driven approach to scenario-based requirements engineering. The approach, which is based on Computer Automated Multi-Paradigm Modeling (CAMPaM), aims to improve the software process. A framework is given and implemented to reason about models of systems at multiple levels of abstraction, to transform between models in different formalisms, and to provide and evolve modeling formalisms.

Relevance to adapID

The adapID project aims to design advanced (privacy-friendly) applications that use the Belgian e-ID card. These applications have anonymity, control, and security requirements. Hence, the methodologies presented in this chapter can help the application designers to address these constraints early on in the software development, and to predict or validate whether the requirements are satisfied in their design. Using multi-paradigm modelling principles yields several important results as well:

- Multiple levels of abstraction allow to control the complexity of the design process. Implementation details are hidden in the conceptual design phase. Hence, the designer does not need to be a security or privacy expert to build controlled anonymous applications using credentials.

- Multiple models are also introduced at the same level of abstraction. Each model exposes a different view on the application.

- Appropriate transformation rules are defined. The rules allow to generate a different view on the application automatically.

8.2 A Modelling and Simulation Based Process for Dependable Systems Design

- Miriam Zia, Sadaf Mustafiz, Hans Vangheluwe and Jörg Kienzle. A modelling and simulation based process for dependable systems design. *Journal on Software & System Modeling*, 2007.

Summary

Dependability can be improved by making use of fault tolerance techniques. The de-facto example in the real-time system literature of a pump control system in a mining environment is used to demonstrate our model-based approach. In particular, the system is modelled using the Discrete EVent system Specification (DEVS) formalism, and then extended to incorporate fault tolerance mechanisms. The modularity of the DEVS formalism facilitates this extension. The simulation demonstrates that the employed fault tolerance techniques are effective. That is, the system performs satisfactorily despite the presence of faults. This approach also makes it possible to make an informed choice between different fault tolerance techniques. Performance metrics are used to measure the reliability and safety of the system, and to evaluate the dependability achieved by the design. In our model-based development process, modelling, simulation and eventual deployment of the system are seamlessly integrated.

Background

Dependable systems must satisfy a set of functional requirements, and in addition, must adhere to constraints which ensure correct behaviour of the system. Safety, security and reliability are a few such dependability requirements. The necessity to satisfy these constraints has spawned new fields of research. The most prominent area is that of fault tolerant systems, and the introduction of fault tolerance design in the software development process is an emerging topic. Research has been carried out in *formal modelling* and analysis of fault tolerance properties [PvH04, BACP96], using either natural language description of models, probabilistic models, figures of fault-trees or Markov models.

Contributions

Although there have been frameworks developped to aid in the production of dependable systems, to our knowledge, current approaches do not offer a modelling and simulation based process such as the one proposed here.

We developed the modelling and simulation based process illustrated in Fig. 8.1 for designing a dependable system. All steps in the evolution, from

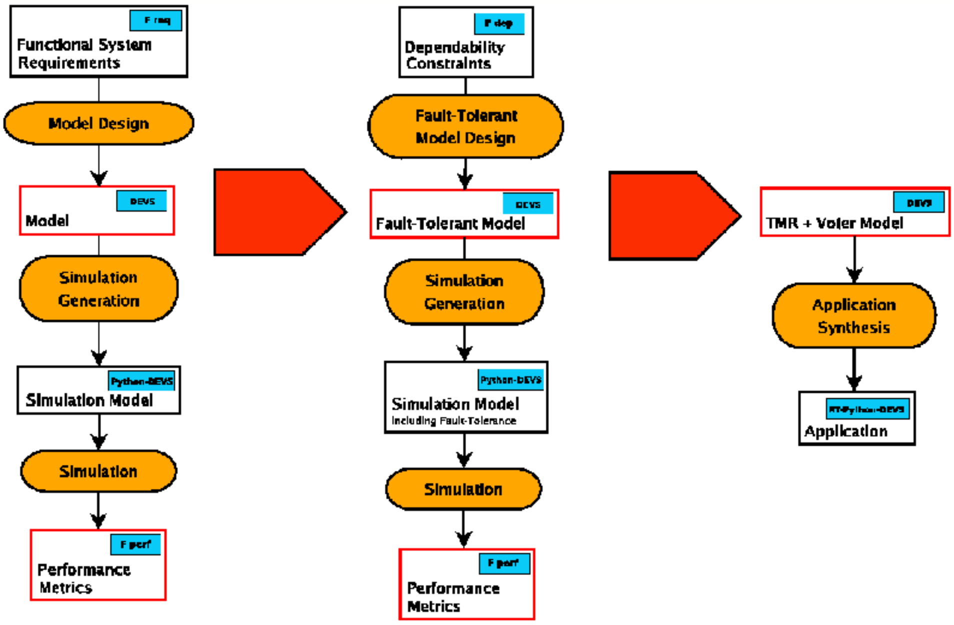


Figure 8.1: The Model-based Process

initial requirements and constraints to final system, are explicitly modelled. Models at various stages of the process are each expressed in the most appropriate formalism. Transformations themselves are also modelled explicitly, so no knowledge is left implicit.

The process allows us to predict the behaviour of a specific system, and compare it to the behaviour of a fault-tolerant implementation of the same system. This is done through a sequence of manual activities. First, from functional requirements, a model is derived which represents the structure of a chosen system. A fault injection mechanism is also modelled as a means to generate faulty behaviour of the system. Simulation results indicate how the system performs in the presence of faults, and whether it conforms to the specified requirements. Secondly, from dependability constraints, a fault-tolerant model is created which includes techniques designed to improve on the initial system. A fault-tolerant simulation model is derived and simulated to gather performance data. This data reflects the dependability constraints that must be satisfied by the system. In this paper, system models are constructed in the DEVS formalism, and simulation models are developed using PythonDEVS [BV01]. Finally, once performance metrics indicate that dependability constraints are satisfied, the fault injection mechanism is separated from the fault-tolerant model, and the final application can then be synthesized using Real-time PythonDEVS [Bor03].

8.3 Model-Driven Design and The Application to Dependability Analysis

- Sadaf Mustafiz, Ximeng Sun, Jörg Kienzle, Hans Vangheluwe. Model-Driven Assessment of Use Cases for Dependable Systems. In *MoDELS*, pages 558-573, 2006.
- Sadaf Mustafiz, Ximeng Sun, Jörg Kienzle, Hans Vangheluwe. Model-Driven Requirements Assessment for Dependable Systems. In *Journal on Software & System Modeling*, 2007.

Summary

We use a probabilistic extension of statecharts to model the system requirements. The model is then evaluated analytically based on the success and failure probabilities of events. The analysis may lead to further refinement of the use cases by introducing detection and recovery measures to ensure dependable system interaction. A visual modelling environment for our extended statecharts formalism supporting automatic probability analysis has been implemented in AToM^3, A Tool for Multi-formalism and Meta-Modelling. Our approach is illustrated with an elevator control system case study.

Background

Complex computer systems are increasingly built for highly critical tasks from military and aerospace domains to industrial and commercial areas. Failures of such systems may have severe consequences ranging from loss of business opportunities, physical damage, to loss of human lives. Systems with such responsibilities should be highly *dependable*.

On the software developer's part, this involves acknowledging that many exceptional situations may arise during the execution of an application, and providing measures to handle such situations to maintain system reliability and safety. Any such exception that is not identified during requirements elicitation might potentially lead to an incomplete system specification during analysis, and ultimately to an implementation that behaves in an unreliable way.

Contributions

For the purpose of analysis, we introduce probabilities in use cases. The value associated to each interaction step represents the probability with which the step succeeds. Our proposed process is illustrated in Fig. 8.2.

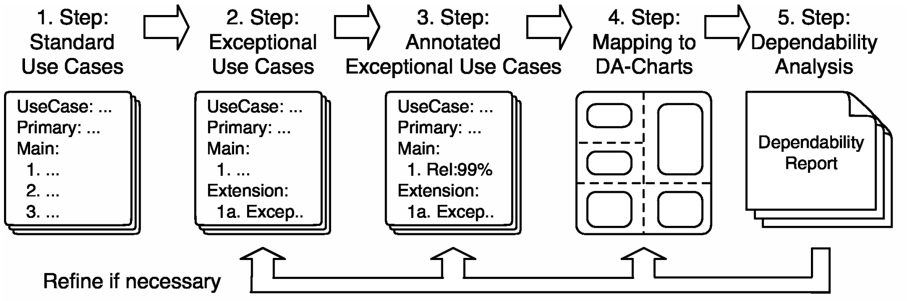


Figure 8.2: Model-Driven Process for Assessment and Refinement of Use Cases

First, the analyst starts off with standard use case-driven requirements elicitation (step 1). Using the exceptional use case approach described in [SMKD05] the analyst discovers exceptional situations, adds detection hardware to the system if needed, and refines the use cases (step 2). Then, each use case step that represents an interaction with a secondary actor is annotated with a probability value that specifies the chances of success of the interaction (step 3). Additionally, each interaction step is annotated with a safety tag if the failure of that step threatens the safety of the system. Next, each use case is mapped to a DA-Chart (step 4). The DA-Charts are then mathematically analyzed by our dependability assessment tool (step 5) and a report is produced. Steps 3, 4, and 5 are the main contributions of this work.

The assessment report allows the analyst to decide if the current system specification achieves the desired reliability and safety.

8.4 A Methodology for Anonymity Control in Electronic Services using Credentials

- Vincent Naessens, *A methodology for anonymity control in electronic services using credentials*, Phd, Department of Computer Science, K.U.Leuven, Leuven, Belgium, June 2006, ISBN:90-5682-711-1. URL: `http://www.cs.kuleuven.be/publicaties/doctoraten/cw/CW2006\_03.abs.html`

- Vincent Naessens and Bart De Decker, A methodology for designing controlled anonymous applications, In *Security and Privacy in Dynamic Environments*, volume 201/2006 of *IFIP International Federation for Information Processing*, (Fischer-Hübner, S., Rannenberg, K., Yngström, L. and Lindskog, S., eds.), pages 111–122. IFIP TC11, Springer Boston, 2006. DOI: 10.1007/0-387-33406-8.

Summary

This work presents a methodology for designing controlled anonymous applications using the multi-paradigm modelling principles. We thereby mainly focus on anonymity control enforcement at the application-level. Anonymous credentials are used as a basic block. The methodology aims to control the complexity of the development process and tries to achieve a reasonable trade-off between the requirements of all stakeholders in the system. The methodology consists of three phases: an analysis phase, a conceptual design phase and an implementation phase.

Multiple formalisms are introduced at the conceptual design phase, namely the flow chart formalism, a Petri Net formalism and a linkability graph formalism. Moreover, a set of transformation rules are defined to transform a model within one formalism to a model within another formalism. The advantages are twofold. First, this strategy allows to make multiple levels of abstraction (Flow Chart versus Petri Net) during the conceptual design phase. Second, it allows to return multiple views to the designer at the same level of abstraction.

Performance properties are derived from the Petri Net; anonymity/trust properties are derived from the linkability graph. The methodology is further fine-tuned. Multiple metrics are defined to measure performance properties and anonymity/trust properties.

Finally, conflict resolution strategies are described. An anonymous credential system is used as basic block to realize the conceptual design. The conceptual design is transformed to high level credential primitives. A linkability graph returns the provable links between the basic blocks at the implementation level. The pieces of evidence that must be stored by service providers to enable control measures can be derived from this graph. We further address the consistency between the linkability graph at the conceptual level and the linkability graph at the implementation level.

Background

A number of privacy-enhancing technologies have been developed in the last years in order to make systems that comply with the privacy requirements. Some of these technologies can be used as building blocks for secure systems. However, designing controlled anonymous applications is no sinecure. An important reason is that this class of applications deals with opposite requirements (anonymity requirements versus control requirements). Moreover, adding building blocks for anonymity control has an impact on other non-functional requirements. Third, the complexity of e-systems increases. Finally, building blocks for anonymity control are often only accessible to security experts.

Contributions

This work is situated at the border between security/ anonymity and software engineering.

In the security area, this work focusses on the design of controlled anonymous applications using anonymous credentials. The emphasis lays on anonymity control enforcement at the application level. The exploration of this domain yields several important results:

- The design of privacy-enhancing applications is driven by the requirements of all stakeholders.
- Anonymous credentials are a crucial building block for anonymity control. However, using anonymous credentials may cause a substantial overhead to actions. Therefore, reasoning about the performance properties during the conceptual design phase cannot be underestimated.
- Trust is inevitable when offering conditional anonymity to users. However, as conditional links can only be revealed by external entities, the level of trust that users must have in service providers for privacy protection is minimal.
- The impact of deanonymization paths on anonymity/trust/performance properties and selected acceptable deanonymization paths with regard to anonymity/performance/trust requirements is addressed.
- Anonymity properties (i.e., unlinkabilities) and accountability properties (i.e., provable links) can be derived from a linkability graph which is extended with conditional/ provable links.

8.5 Addressing Security Requirements through Multi-formalism Modelling and Model Transformation

- Miriam Zia, Ernesto Posse, Hans Vangheluwe. Verification of Security Requirements through Multi-formalism Modelling and Model Transformation. In *2nd International Conference on Software and Data Technologies*. ICSOFT 2007. Submitted.

Summary

Our methodology supports the verification of security properties using the model checker FDR2 on CSP (Communicating Sequential Processes) models. This low-level constraint checking is performed through model refinements, from a behavioural description of a system in the Statecharts formalism. The contribution of this paper lies in the combination of various formalisms and transformations between them. In particular, mapping Statecharts onto CSP models allows for combination of the deterministic system model with non-deterministic models of a system's environment (including, for example, possible user attacks). The combination of system and environment models is used for model checking. To bridge the gap between these Statechart and CSP models, we introduce kiltera, an intermediate language that defines the system in terms of interacting processes. kiltera allows for simulation, real-time execution, as well as translation into CSP models. An e-Health application is used to demonstrate our approach.

Background

Whereas the field of domain-specific security has not yet been thoroughly investigated, research has been conducted in the study of computer security and electronic privacy, and has mostly addressed network and communication applications. Work has been carried out in representing security threats in terms of attack graphs [SHJ$^+$02] and in developing model-checking methods for the verification of the security of communication [Low96] or registration protocols [GL97]. These approaches require detailed models of the protocols, described in languages such as the Communicating Sequential Processes (CSP) [Hoa85] or LOTOS [ISO89], to be checked with verficiation tools, such as the Failures Divergences Refinement Checker (FDR2). What these approaches do not offer, however, is a general methodology for representing the problem at a high level, in such a way as to abstract complex details from, for example, a designer who needs to check that certain properties hold on a new security protocol being developped.

Contributions

In [YLWD05] and [RW06], transformations are defined, which translate high-level behavioural models of a system into models suitable for verification with FDR2. This work represents a step towards a general methodology for the verification of requirements through multi-formalism modelling and model transformation. An example of such a methodology has been proposed for designing controlled anonymous applications [Nae06]. Our contribution lies in introducing a model-driven methodology geared towards security constraint analysis and verification.

We present such a process for security requirement analysis. As depicted in Figure 8.3, this process allows for low-level constraint checking, through model refinements, from higher-level system specifications. At a higher level of abstraction, we have a behavioural description of the system in a known formalism, such as in Statecharts. At a lower level, verification of desired

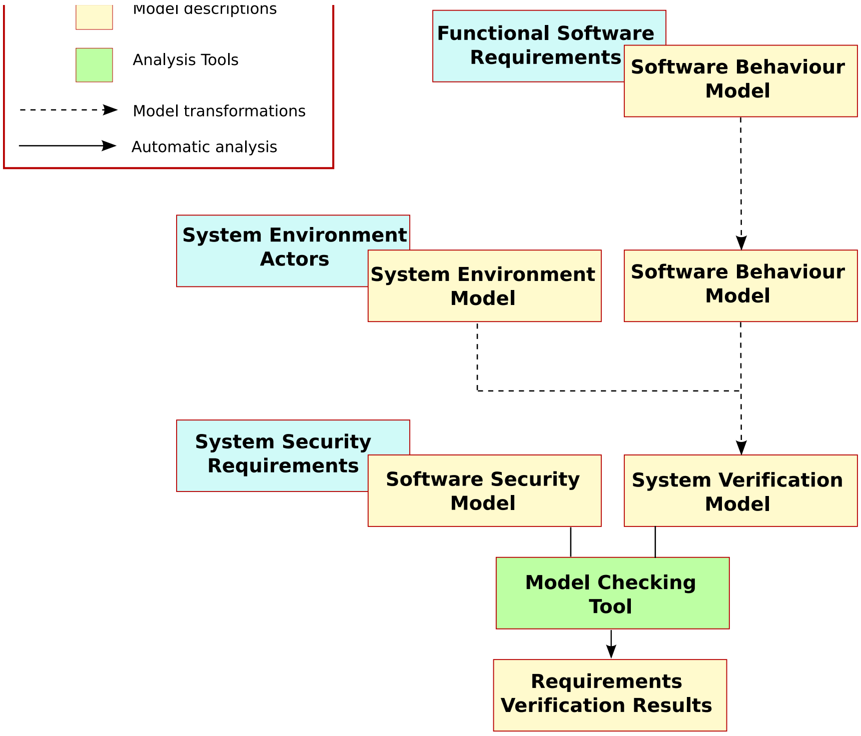


Figure 8.3: A model-based process for security requirement analysis.

properties is performed with existing tools, such as FDR2, on CSP models. Since models in Statecharts only depict required behaviour but lack the detail of the environment actors on this system, our approach introduces a middle step in the design, which acts as a central hub for model transformations. Although Statechart models could be verified at this level, they are first translated into a language which allows for the concise description of non-deterministic behaviour. These intermediate models are described in the modelling language kiltera [PV07]. In turn, the process view of the kiltera language facilitates a straightforward translation into a CSP model, which is subsequently checked to ensure that, given the actors involved in the environment, the original system design adheres to the required security constraints.

8.6 kiltera: a Simulation Language for Timed, Dynamic-structure Systems

- Ernesto Posse and Hans Vangheluwe. kiltera: a Simulation Language for Timed, Dynamic-structure Systems. In *Proceedings of The 40th Annual Simulation Symposium*, 2007.

Summary

This paper introduces a simulation language for dynamic systems with a explicit notions of time and structural change. The language is presented first informally. Then, a formal semantics in terms of *timed labelled-transition systems* is provided. Finally, the paper discusses an application to the modelling and simulation of an adaptive network architecture, where jobs are dispatched to a set of servers grouped in "nodes." The structure of the network adapts itself to deal with load-balancing, according to time constraints.

Background

Modelling systems with dynamic structure has been around since the introduction of the Actors model by Hewitt in 1973 [CHS73] and further developed by Agha [Agh86]. At the same time multiple formalisms and languages have been proposed that have an explicit notion of time, for instance Zeigler's DEVS formalism [ZPK76] or Reed and Roscoe's Timed CSP language [Rr88]. Nevertheless, languages that combine both the capability to undergo structural change and an explicit notion of time have not been explored.

Contributions

kiltera combines the interaction and mobility model of the π-calculus [MPW89] with the time model of Timed-CSP [Rr88] in a unique framework. Its formal semantics are given in terms of timed-labelled transition systems, which provides a basis for formal system analysis as well as a framework to relate kiltera to other languages and formalisms. A real-time simulator has been developed. The language provides high-level constructs such as synchronous and asynchronous communication, pattern-matching, list-comprehensions and process arrays, making it adequate to model complex discrete-event dynamic systems.

8.7 A Model-Driven Approach to Scenario-Based Requirements Engineering

- Ximeng Sun. A Model-Driven Approach to Scenario-Based Requirements Engineering. 102 pages. Submitted as a Master's thesis (supervised by Prof. Hans Vangheluwe).

- Ximeng Sun, Hans Vangheluwe. Transforming Software Requirements by Meta-modelling and Graph Transformation. In *Model-Driven Engineering Languages and Systems: 10th International Conference*. MoDELS'07. Submitted.

Summary

The model-driven approach starts with modeling requirements of a system in scenario models and the subsequent automatic transformation to state-based behavior models. Then, either code can be synthesized or models can be further transformed into models with additional information such as explicit timing information or interactions between components. These models, together with the inputs (e.g., queries, performance metrics, test cases, etc.) generated directly from the scenario models, can be used for a variety of purposes, such as verification, analysis, simulation, animation and so on.

A visual modeling environment is built in AToM^3 using *Meta-Modeling* and *Model Transformation*. It supports modeling in Sequence Diagrams, automatic transformation to Statecharts, and automatic generation of requirements text from Sequence Diagrams.

An application of the model-driven approach to the assessment of use cases for dependable systems is shown.

Background

Some benefits can be gained by using scenarios, such as, minimizing the gap between specification and implementation, avoiding mismatches between user/engineer view, providing a rich view of goals, actions and experiences of users and getting good concepts for extending and redesigning existing systems.

While scenarios have become an established technique in the requirements engineering process, many questions still remain for further research. Some of them are listed here:

- Keeping consistency among the models used at different phases (or levels) of the process;

- Keeping consistency of the models when they evolve with requirements;

- Reusing requirements models;

- Deciding when the scenario specification is complete;

- Capturing and analyzing non-functional requirements.

Computer Automated Multi-Paradigm Modeling (CAMPaM) [VdL03] aims to simplify the modeling of complex systems by establishing a framework to reason about models of systems at multiple levels of abstraction, transforming between models in different formalisms, and providing and evolving modeling formalisms. Based on the aspects that it addresses, CAMPaM can be naturally applied to the problem of improving scenario-based requirements engineering. For example, modeling, analysis and validation at multiple levels and in different views, can be achieved by model abstraction and multi-formalism modeling in an automatic and consistent way. As another example, rich methods for verification, analysis, simulation and execution of the target system provided by CAMPaM, can maintain the agreement with customers and increase their acceptability and satisfaction.

Contributions

We have proposed a model-driven approach to scenario-based requirements engineering, as shown in Figure 8.4.

The approach starts at the top level of the figure which models requirements of a system in scenarios by means of formalisms such as Use Cases, Sequence Diagrams, or Use Case Charts [Whi05]. Then, state-based behavior models (e.g., Statecharts [Har87] or DCharts [Hui04] models) can be generated automatically by model transformation. At this point, we can already use generated hierarchical state machines (HSMs) to do some automatic simulation by tools such as SCASP [Whi05], RHAPSODY [HKP05], and SVM [Fen03], and even synthesize code for execution by tools such as the Statechart compiler SCC [Hui04]. However, HSMs are limited to the cases where explicit timing information or interactions between components are important for simulation, analysis and verification. So we further transform these HSMs to models in formalisms such as DEVS, Communicating Sequential Processes (CSP), or Timed Petri Nets (TPN). We can now map HSMs into models of one single formalism, kiltera [PV07], which allows mapping onto other formalisms as shown in Figure 8.4. Since the mapping from HSMs to kiltera is not trivial, we will still need human intervention to aid in refinement.The derived models are used for model checking, analysis, simulation or animation. Another advantage of our approach is that the inputs of these tasks, such as queries, performance metrics and test cases, can be automatically generated from the scenario models done at the beginning. Furthermore, as illustrated in Figure 8.4, the scenario models can be used to generate textual or graphical representation of requirements in a language the customers are familiar with for their evaluation and immediate feedback.

As shown at the top of Figure 8.4, there is a transformation which leads to a dependability analysis model. This shows that it is easy to transform scenario models to some formalism such as DA-Charts for some specific analysis (e.g., dependability analysis) [MSKV06].

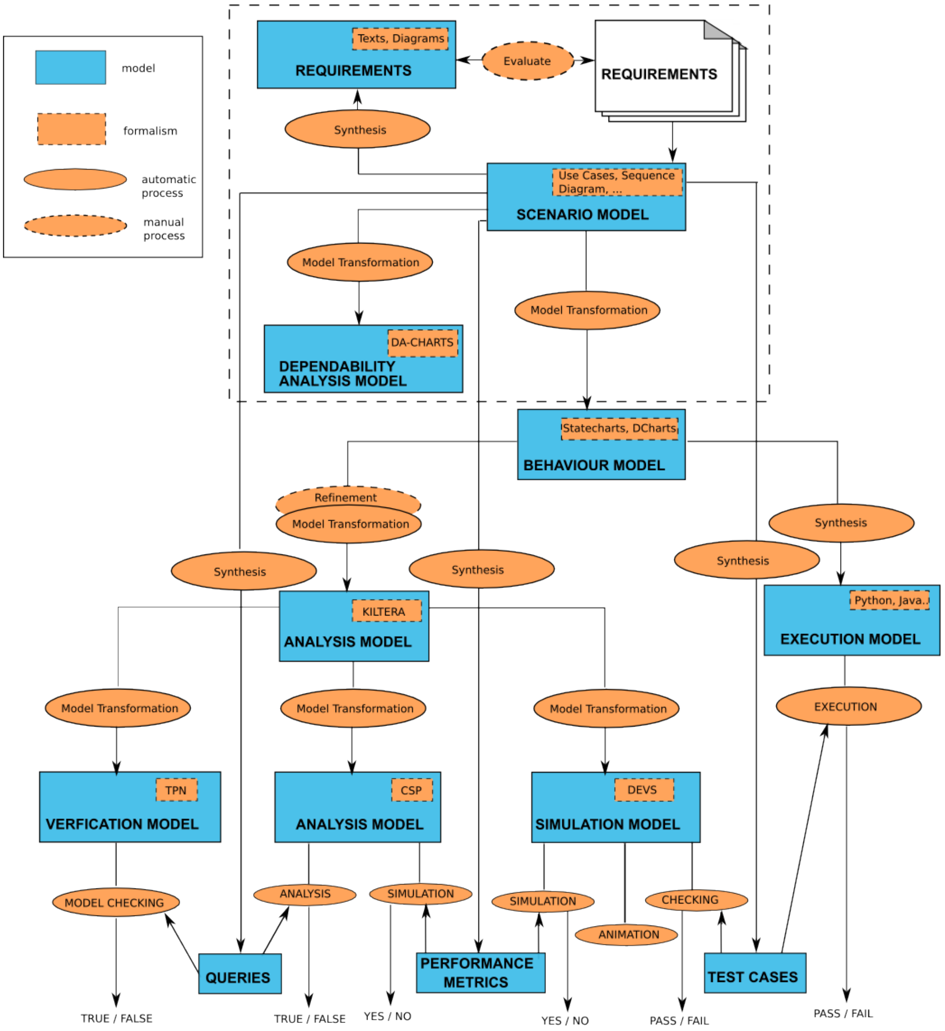


Figure 8.4: A Model-Driven Approach to Scenario-Based Requirements Engineering

Chapter 9

Legal aspects

9.1 Introduction

The roll out of e-ID cards has – to date – not led to a revolution in how people conduct business online or interact with the government. The deployment of advanced e-ID applications should however start the ball rolling. As this development goes on, a number of legal questions crop up.

Can we rely on the market to provide all the necessary e-ID applications and components, with sufficient quality and availability? There may be circumstances, particularly in the e-government context, where the government wants to develop and provide their own applications. All the best intentions to provide citizens with reliable applications, cannot detract from the fact that this might distort the market, in contravention with competition rules and regulation of public sector aid.

Widespread use of e-ID applications begs the question whether we are building the most comprehensive surveillance system in history. Issues of data protection should be considered on two levels. Firstly, the advanced e-ID applications should comply with data protection rules in the way they operate. Secondly, privacy implications of the analysis of the datastreams generated through the use of such applications should be considered. It should be noted that such analysis can be conducted for various purposes, e.g. scientific research, user profiling, identity theft, ...

Obviously, it is quite insufficient for e-ID application providers to simply claim their application can be relied on to protect the user's privacy. On the one hand providers must be able to convince user's of their trustworthiness, on the other user's must be in a position to hold them accountable for breaches of trust. This is where questions regarding evidence come in.

Relevance to ADAPID

The law is often seen as something operating in the background of technological developments, however at times its impact is only all too real. Not only does the law assign consequences to certain design choices, it regulates the conditions within which technology may be deployed. Privacy regulation is a prime example, as it is one of the driving forces behind the search for anonymous and pseudonymous use of an authentication device as the e-ID.

In [DT06] we show that there are still some issues to be resolved to the Belgian e-ID more privacy friendly and the role which could be played by anonymous credentials.

In [Dek05] we look at one of the outputs of advanced e-ID applications, i.e. electronically signed contracts, and look at the rules of evidence that govern them in light of preservation issues.

In [DD07c] we continue the analysis of preservation issues by analysing certain key legal terms and comparing them with concepts from the InterPARES framework on digital archiving.

In [DDD07] we look at logging, which is a basic functionality built into any advanced e-ID application. The rules governing the use of this function in practice are therefor highly relevant.

9.2 The Belgian e-ID Card as an e-Government Tool

- Sven Van Damme and Reshma Thomas. *Special report on e-government: preserving privacy.* In Global Identification, 42-45, October 2006.

Summary

In [DT06] we start by giving an outline of the Belgian e-ID card. The history, the purpose and the functions are described and it is made clear what data can be found on the card. The main focus is on the privacy issues concerning the use of the Belgian e-ID. We recognize that the e-ID can be a highly versatile and valuable instrument, certainly for facilitating eGovernment functions and services, but we point to the fact that some privacy issues are not fully resolved yet. An important principle of the European Data Protection Directive is that personal data must be adequate, relevant and excessive in relation to the purposes for which they are collected or further processed. This principle is also called “data minimization” as it requires that the least possible amount of data be processed about a data subject. This means that the e-ID card should only reveal the information that is necessary for the application. We stress that this is not the case when the Belgian e-ID card is used. The use of certain privacy enhancing technologies may offer a solution to solve this problem. Another problem with the Belgian e-ID is that the National Register number (RRN) can be found four times on the e-ID. We make clear what the benefit is of working with a universal unique identifier, but we point out that it can undermine the citizen's privacy, due to the ability of the identifier to act as a key to uncovering and linking a vast amount of information in order to create a complete personal profile. To finish we discuss the sophisticated solution Austria has come up with to avoid the problems of linkability arising as a result of the use of unique identifiers.

Contributions

The contribution of this paper is that it gives a clear overview of the threats to privacy that arise as a result of the use of the Belgian electronic identity card. It points to the privacy threats posed by e-government systems and more specifically by the use of unique identifiers, as they make it easier to exchange data across several administrative systems. The fact that in the current model of the Belgian e-ID the RRN is in the end user certificate facilitates the use of this number and in this way increases the chance that personal privacy is violated. As a possible solution for this problem we described the Austrian context-dependent identifier. Publicly readable electronic information on an e-ID poses a severe privacy threat: whenever the citizen uses his card, a service provider may read identifying information from the card.

One of the objectives of the ADAPID project is to propose better solutions and to make the future Belgian e-ID more privacy friendly. This paper makes clear that there still are a few issues that have to be solved. This can be achieved when anonymous credentials will be integrated.

9.3 Preservation of Signed Electronic Records

- Hannelore Dekeyser. *Preservation of Signed Electronic Records.* In Proceedings of the DLM Conference Budapest, 8 p., October 5-7 2005, `http://ec.europa.eu/transparency/archival_policy/dlm_forum/proceed2005_en.htm`.

Summary

In [Dek05] a brief overview of the EU legal framework applicable to electronic signatures is given, with reference to the Belgian transposition. The roll out of the electronic identity card gives all Belgian citizens the means to create digital signatures. Therefor a reliable preservation strategy for signed digital records is required. We explore some preservation strategies for digitally signed records, in particular hard copy, emulation and migration. What follows is an illustration of how a Belgian judge might approach the different preservation strategies in a contractual dispute.

Contributions

To date, little to no known jurisprudence is available on the validity of electronic signatures. Unsurprisingly, no know jurisprudence exists regarding long-term preservation of signed documents either. Until jurisprudence on this topic comes into existence, any legal analysis necessarily remains hypothetical. Fair and consistent jurisprudence will not arise without an in depth understanding of the technical issues at play and their legal consequences. This article aims to support such understanding.

Though contracts are only a fraction of the output of advanced e-ID applications, from a legal point of view they are highly significant. Contracts law was one of the first legal domains to accomodate electronic signatures. The rules of evidence that deal with conflicts about contracts are also the default rules of evidence. Thus, the way in which the legal world deals with electronically signed contracts will almost certainly permeate into other domains of law.

9.4 Speaking of record's preservation: avoiding terminological confusion between legal experts and archivists

- Hannelore Dekeyser and Jos Dumortier. *Speaking of record's preservation: avoiding terminological confusion between legal experts and archivists.* Submitted to InterPARES (The International Research on Permanent Authentic Records in Electronic Systems, `http://www.interpares.org`).

Summary

In [DD07c] we look at 3 selected legal texts dealing with digital documents in relation to a number of key concepts of the InterPARES digital archiving framework, specifically 'authenticity', 'identity' and 'integrity' of digital records. The texts in question are the United Nations Convention on the Use of Electronic Communications in International Contracts 2005 (CEUCIC), the European e-Invoicing Directive and the European Directive on Distance Marketing of Financial Services.

Contributions

The analysis shows that the terminology concerning electronic records as used by the legal community is in a state of flux. On top of this, a number of terms are not used consistently to mean the same thing. To make matters worse, a number of terms have one specific meaning in legal doctrine and quite another in the InterPARES framework.

The legal texts cited here have no direct impact on the advanced e-ID applications developed with the Adapid project. However, the terms and concepts used in EU directives and - to a lesser extent perhaps - international conventions eventually find their way into national legislation and jurisprudence. Quite often, novel terms and concepts are reused beyond the scope of the legal text that initially introduced them.

9.5 Logging use of information systems: logfiles in light of C.L.A. nr. 81.

- Hannelore Dekeyser and Sven Van Damme and Jos Dumortier. *Gebruik van informatiesystemen registreren: logbestanden in het licht van C.A.O. nr. 81* (*Logging use of information systems: logfiles in light of C.L.A. nr. 81.*). Submitted to Privacy & Informatie (Privacy & Information journal).

Summary

Employee's activities are routinely registered in the log files of computer applications. The interests of the company clash with the right to privacy of the workforce. C.L.A. nr. 81 intends to resolve this issue by clarifying what the principles of legality, finality and proportionality mean in this context. Unfortunately, C.L.A. nr. 81 isn't very clear, also the focus lies squarely on abuse of e-mail and internet. In [DDD07] we explore the impact of this piece of regulation on logging activities.

Contributions

C.L.A. nr. 81 is applicable to logfiles, but it's provisions can only be applied with difficulty to the registration of what is in principle legitimate use of computer applications. In particular the indirect procedure for identification – which applies by default – is problematic. What is the use of recording an audit trail, if consulting it is prohibited?

Logging is a basic functionality built into any advanced e-ID application. The applicable rules for the use of this function in practice are therefor highly relevant.

9.6 Work in progress

9.6.1 Public sector aid

Summary

The starting point of this research is the question whether the government is allowed to set up a trusted archival service from a competition law point of view, but can of course be extended to other services set up by the government. We conduct research on the possibilities the government has to set up such a service and more specifically we look at the state aid rules in the EC Treaty.

The EC Treaty attempts to achieve a balance between Member States and the Community by laying down a broad prohibition on the granting of State aid which distorts or threatens to distort competition by favouring certain undertakings or the production of certain goods insofar as it effects trade between Member States.

The state can enter the market and take part in the economy either through a body that has separate legal personality or through a body that is integrated into the state administration. For the application of state aid rules this distinction is not relevant. Aid granted to public undertakings, whether or not pursuing public duties is subject to the same rules applying to all forms of aid granted by Member States.

The government can decide to set up an undertaking to deliver these services. If the State is acting in a way that a private investor would in a market economy, for example in providing loans or capital on similar terms to that of a private investor, it is not providing State aid within the meaning of Article 87(1). (i.e. the market economy investor principle). We examine which criteria the European Commission and the European Courts have set out to decide whether an investment is done on market terms.

The government can also decide to impose obligations in the general economic interest upon undertakings. After we have tried to describe how a service of general economic interest can be defined, we look at the case law of the European Court of Justice and that of the Court of First Instance concerning this issue. More specifically we will try to give guidance on the issue of whether or not financial support to undertakings for these obligations could be exempt from state aid rules in the sense that such measures being merely compensation for specific burdens or obligations could not constitute an advantage within the meaning of art. 87 (1).

We also look if the state aid rules in the EC Treaty offer any other possibilities to set up these kinds of services.

Contributions

This research tries to give an answer on the question when the government is allowed to finance a certain public service. The EC Treaty prohibits the granting of state aid when it threatens to distort competition and when trade between Member States is affected, but the case law and practice of the European Commission and the European Courts have shown that in certain cases the financing by the public authorities is not state aid. It is also possible that there is state aid, but aid compatible with the common market. It is our objective to analyse the relevant practice and case law of the European Commission and of the European Courts, to give a clear overview of the possibilities the government has when it wants to set up a certain public service.

One of the applications in the ADAPID project is the long-term storage of digital information. How to preserve, use, and verify over time administrative, commercial or business records, is highly important within the juridical-administrative context. It is possible that private sector companies are prepared to deliver these kinds of services. But if this is not the case, it will be up to the government to take measures. In this context, the regulation on the financing of public services by public authorities becomes highly relevant.

9.6.2 Processing of personal data for scientific research

Summary

Scientific researchers are often confronted with personal data and in that way in principle have to fulfill the obligations that are imposed by the Data Protection Act. This can cause a lot of problems for scientific researchers. One could say that scientific research and privacy law are often in a conflict situation. On top of that, in Belgium, there is a very strict interpretation of what can be considered as personal data. All data which still can be linked to the individual are regarded as personal data even when they are processed by a person who can not make that link. This can be very important for scientific researchers, but also for a lot of e-ID card applications this can be of great significance.

The legislator realized that was difficult for scientific researchers to fulfil all the obligations that can be found in the Data Protection Act. That is why the legislator has been a bit more lenient towards research. A balance had to be found between the interests of the scientific research and the protection of the privacy of the persons whose data are processed for the research.

Contributions

In our research we will point to the different legal impediments scientific researchers are confronted with. We will give an overview of the measures the legislator has taken to reconcile the different interests and try to analyse whether these measures are adequate.

This research can be linked to the advanced applications developed in the ADAPID project on two levels. The first level is aptly illustrated by one of the use cases in the ADAPID requirements study, more specifically the IRIS-Qubec case, which dealt with the processing of personal data. It became clear that the e-ID can play an important role in ensuring that the personal data are only processed for scientific or historical purposes, an important requirement of privacy law. The second level is the case where the data generated through use of the e-ID applications in turn becomes the object of scientific analysis. An example would be empirical research on the use of the e-ID in e-government applications to test the thesis of unlinkability. This research is a prime example of the difficulties caused by the expansive definition of the term personal data as held by the Belgian Privacy Commission. Difficulties arise both during use of the different e-ID applications, as in scientific analysis about their use.

9.6.3 Trust and accountability

Results of this research are intended to be incorporated in the PhD of Hannelore Dekeyser (Title: The Legal Value of Digital Records. A study of the legal issues arising out of long-term archiving of digital records.)

The issue of trust is researched in connection with accountability. Trust research is an increasingly popular research domain in computer sciences, specifically in the area of intelligent systems. However, it is not self-evident what is meant precisely with the concept 'trust' in this context. So much is certain, the term stems from the realm of human relationships and was then used metaphorically in the context of computer systems and human-computer interactions. Considering the babylonic confusing surrounding the term trust in the human sciences (psychology, sociology, economics,...) this is not an easy task. In short, the research asks whether we can truly trust a computer system (or computer systems each other), by answering this question we hope to learn more about the appropriate approach to accountability (who/what is accountable)

Chapter 10

Conclusions

We have presented in this report the main research results obtained by the ADAPID project partners in the first 20 months of the project.

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 sought for in a credential system, together with a number of possible application domains. We have then presented a number of research papers we have either published and submitted for publication.

The area of privacy-preserving digital credentials is a wide and challenging area of research, with a multitude of interesting questions that still need to be answered. We are currently working on a number of issues, among which we note the development of privacy-enhanced databases to guard against non-consensual disclosures of personal data from central databases. We are also conducting an in-depth comparative survey of the state of the art in privacy preserving digital credentials.

In Chapter 3, we have presented our work on models to measure anonymity. We have first presented a paper that recapitulates the existing models to measure anonymity in communication systems, and explains information-theoretic anonymity metrics, which are based on Shannon entropy. In the second paper, we have explained that often, entropy based anonymity metrics have been used for the evaluation of mixes, without taking into account any user profile information the attacker could have. We have addressed the confusion between the adversary's uncertainty (which is represented by the "entropy of a conditional probability distribution") and the "conditional entropy", which 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 have explained the relationship between the adversary's uncertainty, which is calculated from a particular observation of the anonymity system, 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).

Our work on anonymous communication has been presented in Chapter 4. The first paper presented in this chapter is a survey on anonymous communication, that examines the approaches that have been followed in order to design and implement systems that anonymize the communication layer. In the second paper, we have proposed an anonymous communication infrastructure that enables anonymity revocation is cases of abuse, and that combines several technologies such as digital credentials or key traitor tracing schemes. Our third paper describes a new attack on mix-based anonymous communication, that considers more sophisticated user models than previous work. In particular, we have found that anonymous email services that support replies are even more vulnerable to statistical disclosure attacks.

There are still many open problems in the field of anonymous communications, and we will continue to do research in this field. Two of the topics we intend to investigate in the near future are peer-to-peer anonymous communication networks for low-latency voice communications (voice over IP), and user behavioral models for anonymous email usage.

In our research in the field of secure storage, we have improved privacy enhancing properties of archives. We have demonstrated that the state of the art in Steganographic file systems (file systems that allow to conceal data in such a way that an adversary cannot coerce the users into revealing data) is not suitable when facing continuous surveillance of the system. We have proposed a new scheme that could be used in this scenario. We have also developed an approach for privacy enhanced search in databases, where the owner of the database does not learn about the information retrieved by a user, or even about the content of the query.

Our ongoing research focuses on the third of the ADAPID use cases, the Long Term Archive. Our focus is on the development of a protocol that allows users to store digitally signed documents in an archive, capable of guaranteeing the validity of the signatures over time, as well as the integrity of the archived documents. More research needs to be done in order to push other security properties into the future, such as confidentiality or availability (in the sense of overcoming the obsolescence of software).

In Chapter 6 we present our research on protocols for distribution of trust. We start the chapter with a summary of a paper on some relation between verifiable secret sharing and a class of error-correcting codes, in which we try to shed a new insight on verifiable secret sharing schemes. The second paper presents an overview of some approaches for building secret sharing schemes based on well studied combinatorial objects and shows the relations between them. Finally, we consider multi-party computation in a hierarchical setting, where participants have different capabilities depending on their position in the hierarchy. We obtain a new class of strongly multiplicative linear secret sharing schemes with explicit ideal constructions.

Policies are essential in privacy friendly applications. Several policy languages (e.g. P3P and APPEL) have been extended to support the use of private (anonymous) credentials. A secure and flexible DRM framework has been proposed, and the application of DRM techniques to protect personal data and enforce policies is being studied.

The ADAPID-framework will be extended to support privacy, disclosure and access policies. Also, sticky policies (i.e. policies that are inextricably bound up with the resources they protect) will be considered. Special tamperproof hardware or DRM techniques might be helpful to realize this kind of policies. DRM techniques could also help to better protect sensitive personal data (e.g. e-health records).

Some of the applications considered in the ADAPID project may have to deal with disputes. Evidence to prove or refute dispute claims need to be kept by the applications and provided to an arbiter on request. In this research topic, the selection of the proper evidence material and (semi-) automated dispute handling will be studied.

In Chapter 8, we discussed model-based design processes that can be applied to the implementation of secure, anonymous and privacy-preserving e-ID applications. We have demonstrated a continuous workflow, from system requirements to deployed system, through which dependability constraints are addressed and verified, either through simulation or model checking. First, we presented a methodology for designing controlled anonymous environments, which compromised between privacy requirements and control requirements. Then, we introduced a modeling and simulation based approach which allowed for the analysis and prediction of dependability constraints. Performance metrics derived from simulation results helped in validating whether the system reached satisfactory dependability levels.

In order to address dependability at a higher-level, and earlier on in the design, we proposed on the one hand, a model-based approach to scenario-based requirements engineering, and on the other, an approach for assessing and refining use cases that ensured that the specified functionality met the dependability requirements of the system as defined by the stakeholders.

Finally, a multi-formalism process was described which was geared towards security constraint analysis and verification from initial system requirements, through a series of model transformations.

The legal research has covered four main topics. Firstly, we have looked at issues of data processing and privacy protection regarding the e-ID and advanced e-ID applications in general. Specific attention was given to logging functionality, as well as rules governing analysis of log files and other personal data for scientific research. A second research topic was preservation of evidence, in particular one of the outputs of advanced e-ID applications, namely electronically signed records. The rules of evidence regarding contracts were looked at in particular, since these rules serve as default in other areas. Also, we analyzed certain key legal terms and compared them with concepts from the InterPARES framework on digital archiving. Thirdly, the regulatory framework regarding public sector aid was explored. Past experience has shown that the market may not provide the necessary building blocks – or not provide them with the required quality – for the development of advanced e-ID applications. The government may feel the need to enter the market themselves, thus raising questions regarding the regulations governing the financing of public services by public authorities. Last, but not least, the topic of trust and accountability has been the subject of research. Though this research is highly conceptual in nature, it aims to clarify which approach to take regarding shared accountability between systems, users and operators.

On the first two topics, a number of results can be reported. Where data protection is concerned, we have formulated insights into the privacy challenges facing the Belgian e-ID and the improvements that could be made. Also we have gained an understanding of how data protection rules affect logging activities. With regard to preservation of evidence, we have formulated the challenges posed to preservation of signed electronic records and some avenues for a legal solution. Also, a conceptual analysis was made of key legal terms with respect to electronic records, leading to an insight on the pitfalls and opportunities for the application of digital archiving principles.

Results on the other two topics are still pending, however considerable progress has been made to date. Further work is planned on all four topics. The work on public sector aid will be finalized in the coming months. The work on data protection will continue throughout the project, since this underpins the advanced e-ID applications. Likewise, the work on preservation of evidence and on trust and accountability will continue.

Bibliography

[A P] A p3p preference exchange language 1.0 (appel1.0). URL: http://www.w3.org/TR/P3P-preferences/.

[ACJT00] Giuseppe Ateniese, Jan Camenisch, Marc Joye, and Gene Tsudik. A practical and provably secure coalition-resistant group signature scheme. In *CRYPTO '00: Proceedings of the 20th Annual International Cryptology Conference on Advances in Cryptology*, volume LNCS 1880, pages 255–270. Springer-Verlag, 2000.

[Agh86] Gul A. Agha. *Actors: A Model of Concurrent Computation in Distributed Systems.* MIT Press, 1986.

[AK03] Dakshi Agrawal and Dogan Kesdogan. Measuring anonymity: The disclosure attack. *IEEE Security & Privacy*, 1(6):27–34, 2003.

[BACP96] J. Boue, J. Arlat, Y. Crouzet, and P. Petillon. Verification of fault tolerance by means of fault injection into VHDL simulation models. Technical report, LAAS-CNRS, 1996.

[BBS04] Dan Boneh, Xavier Boyen, and Hovav Shacham. Short group signatures. In *CRYPTO '04: Proceedings of the 24th Annual International Cryptology Conference on Advances in Cryptology*, volume LNCS 3152, pages 41–55. Springer-Verlag, 2004.

[BDDD06] Stefan Brands, Liesje Demuynck, and Bart De Decker. A practical system for globally revoking the unlinkable pseudonyms of unknown users. Report CW 472, K.U.Leuven, Department of Computer Science, December 2006. URL: http://www.cs.kuleuven.be/publicaties/rapporten/cw/CW472.abs.html.

[BDDD07] Stefan Brands, Liesje Demuynck, and Bart De Decker. A practical system for globally revoking the unlinkable pseudonyms of unknown users. In *Australasian Conference on Information Security and Privacy*, 2007. *accepted.*

[Bor03] Spencer Borland. Transforming statechart models to DEVS. Master's thesis, School of Computer Science, 2003.

[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.

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

[Bri89] Ernest F. Brickell. Some ideal secret sharing schemes. In *Advances in Cryptology - EUROCRYPT '89*, volume 434 of *Lecture Notes in Computer Science*, pages 468–475. Springer, 1989.

[BV01] Jean-Sebastien Bolduc and Hans L. Vangheluwe. The modelling and simulation package PythonDEVS for classical hierarchical DEVS. MSDL technical report MSDL-TR-2001-01, McGill University, 2001.

[CDG$^+$05] Ronald Cramer, Vanesa Daza, Ignacio Gracia, Jorge Jiménez Urroz, Gregor Leander, Jaume Martí-Farré, and Carles Padró. On codes, matroids and secure multi-party computation from linear secret sharing schemes. In *Advances in Cryptology - CRYPTO 2005*, volume 3621 of *Lecture Notes in Computer Science*, pages 327–343. Springer, 2005.

[Cha81] David Chaum. Untraceable electronic mail, return addresses, and digital pseudonyms. *Communications of the ACM*, 24(2):84–88, 1981.

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

[Cha83] David Chaum. Blind signature systems. In *Proceedings of CRYPTO'83*, pages 153–156. Plenum Press, 1983.

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

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

[CHK$^+$06] Jan Camenisch, Susan Hohenberger, Markulf Kohlweiss, Anna Lysyanskaya, and Mira Meyerovich. How to win the clonewars:

efficient periodic n-times anonymous authentication. In *Proceedings of the 13th ACM Conference on Computer and Communications Security*, pages 201–210. ACM, 2006.

[CHL05] Jan Camenisch, Susan Hohenberger, and Anna Lysyanskaya. Compact e-cash. In *Proceedings of EUROCRYPT'05*, volume LNCS 3494, pages 302–321. Springer-Verlag, 2005.

[CHS73] Peter Bishop Carl Hewitt and Richard Steiger. A universal modular actor formalism for artificial intelligence. In *Proceedings of IJCAI 1973*, 1973.

[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. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. In *Proceedings of EUROCRYPT '01*, volume LNCS 2045, pages 93–118. Springer-Verlag, 2001.

[CL02] Jan Camenisch and Anna Lysyanskaya. A signature scheme with efficient protocols. In *In International Conference on Security in Communication Networks (SCN)*, volume LNCS 2576, pages 268–289. Springer-Verlag, 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 '04: Proceedings of the 24th Annual International Cryptology Conference on Advances in Cryptology*, volume LNCS 3152, pages 56–72. Springer-Verlag, 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.

[Cot] Lance Cottrell. Mixmaster & remailer attacks. Unpublished manuscript, `http://www.obscura.com/~loki/remailer/remailer-essay.html`.

[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.

[CS97] Jan Camenisch and Markus Stadler. Efficient group signature schemes for large groups (extended abstract). In *CRYPTO '97: Proceedings of the 17th Annual International Cryptology Conference on Advances in Cryptology*, volume LNCS 1296, pages 410–424. Springer-Verlag, 1997.

[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 *EUROCRYPT'91*, volume LNCS 547, pages 257–265. Springer-Verlag, 1991.

[Dam90] Ivan Damgard. Payment systems and credential mechanism with provable security against abuse by individuals. In *In Proceedings of CRYPTO '90*, volume LNCS 403, pages 328–335. Springer-Verlag, 1990.

[Dan03] George Danezis. Statistical disclosure attacks. In *Proceedings of SEC'03*, volume 250 of *IFIP Conference Proceedings*, pages 421–426. Kluwer, 2003.

[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.

[DD06] George Danezis and Claudia Diaz. Improving the decoding efficiency of private search. In *In Anonymous Communications and its Applications - Proceedings of Dagstuhl Seminar*, page 11, 2006.

[DD07a] George Danezis and Claudia Diaz. Space-efficient private search. In *Financial Cryptography and Data Security - Proceedings of FC'07*, page 15. Springer-Verlag, 2007.

[DD07b] George Danezis and Claudia Diaz. A survey of anonymous communication channels. Submitted to the Journal of Privacy Technology, 2007.

[DD07c] Hannelore Dekeyser and Jos Dumortier. Speaking of record's preservation: avoiding terminological confusion between legal experts and archivists. submitted to InterPARES (The International Research on Permanent Authentic Records in Electronic Systems, 2007.

[DDD05a] Liesje Demuynck and Bart De Decker. Anonymous updating of credentials. Report CW 430, Department of Computer Science, K.U.Leuven, Leuven, Belgium, December 2005. URL: http://www.cs.kuleuven.ac.be/publicaties/rapporten/cw/CW430.abs.html.

[DDD05b] Liesje Demuynck and Bart De Decker. Privacy-preserving electronic health records. In *Communications and Multimedia Security*, volume 3677 of *Lecture Notes in Computer Science*, pages 150–159. Springer, 2005.

[DDD06a] Liesje Demuynck and Bart De Decker. Credential-based systems for the anonymous delegation of rights. Report CW 468, K.U.Leuven, Department of Computer Science, November 2006. URL: http://www.cs.kuleuven.be/publicaties/rapporten/cw/CW468.abs.html.

[DDD06b] Liesje Demuynck and Bart De Decker. How to prove list membership in logarithmic time. Report CW 470, K.U.Leuven, Department of Computer Science, December 2006. URL: http://www.cs.kuleuven.be/publicaties/rapporten/cw/CW470.abs.html.

[DDD07] Hannelore Dekeyser, Sven Van Damme, and Jos Dumortier. Gebruik van informatiesystemen registreren: logbestanden in het licht van c.a.o. nr. 81. (*Logging use of information systems: logfiles in light of C.L.A. nr. 81.*), submitted to Privacy & Informatie (Privacy & Information journal)., 2007.

[DDDJ07] Liesje Demuynck, Bart De Decker, and Wouter Joosen. A credential-based system for the anonymous delegation of rights. In *New Approaches for Security, Privacy and Trust in Complex Environments*, IFIP. IFIP TC11, Springer, May 2007.

[DDM03] George Danezis, Roger Dingledine, and Nick Mathewson. Mixminion: Design of a Type III Anonymous Remailer Protocol. In *IEEE Symposium on Security and Privacy*, 2003.

[DDP06] Ivan Damgard, Kasper Dupont, and Michael Ostergaard Pedersen. Unclonable group identification. Cryptology ePrint Archive, Report 2005/170, 2006.

[DDT07] George Danezis, Claudia Diaz, and Carmela Troncoso. Two-sided statistical disclosure attack. Submitted to Privacy Enhancing Technologies 2007 (PET'07), 2007.

[Dek05] Hannelore Dekeyser. Preservation of signed electronic records. In *Electronic Records Supporting e-Government and Digital Archives*, Budapest, 5-7 October 2005. DLM Conference.

[DFT07] Claudia Diaz, Peter Fairbrother, and Carmela Troncoso. Observable steganographic file systems using pool mixes and universal re-encryption. Technical report, COSIC, ESAT, KU Leuven, February 2007.

[Dia06] Claudia Diaz. Anonymity metrics revisited. In Anonymous Communication and its Applications, number 05411 in Dagstuhl Seminar Proceedings. Internationales Begegnungs- und Forschungszentrum fur Informatik (IBFI), Schloss Dagstuhl, 2006.

[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.

[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.

[DP06] Claudia Diaz and Bart Preneel. Accountable anonymous communication. To appear as Chapter of Security, Privacy and Trust in Modern Data Management, Springer (in print), 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.

[DS04] George Danezis and Andrei Serjantov. Statistical disclosure or intersection attacks on anonymity systems. In *Information Hiding*, volume LNCS 3200, pages 293–308. Springer-Verlag, 2004.

[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.

[DSD04] Claudia Diaz, Len Sassaman, and Evelyne Dewitte. Comparison between two practical mix designs. In *Proceedings of 9th European Symposium on Research in Computer Security (ESORICS'04)*, pages 141–159. Springer-Verlag, LNCS 3193, 2004.

[DT06] Sven Van Damme and Reshma Thomas. Special report on e-government: preserving privacy. *Global Identification*, pages 42–45, October 2006.

[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.

[Fai04] Peter Fairbrother. An improved construction for universal re-encryption. In *Proceedings of Privacy Enhancing Technologies workshop (PET'04)*, volume 3424 of *LNCS*. Springer-Verlag, 2004.

[Fen03] Thomas Huining Feng. An extended semantics for a Statechart Virtual Machine. In A. Bruzzone and M. Itmi, editors, *Summer Computer Simulation Conference. Student Workshop*, pages S147 – S166. Summer Computer Simulation Conference, July 2003. Montréal, Canada.

[GDD06a] Steven Gevers and Bart De Decker. Automating privacy friendly information disclosure. Report CW 441, K.U.Leuven, Department of Computer Science, April 2006. URL: http://www.cs.kuleuven.be/publicaties/rapporten/cw/CW441.abs.html.

[GDD06b] Steven Gevers and Bart De Decker. Privacy friendly information disclosure. In *On the Move to Meaningful Internet Systems 2006: OTM 2006 Workshops*, volume 4277 of *Lecture Notes in Computer Science*, pages 636–646. Springer, 2006.

[GDD07] Steven Gevers and Bart De Decker. Enhancing privacy in identity systems. In *7th Workshop on Privacy Enhancing Technologies*, 2007. *submitted.*

[GL97] F. Germeau and G. Leduc. Model-based design and verification of security protocols using lotos, 1997.

[Goo] Google. `http://www.google.com/`.

[Gor98] D. M. Gordon. A survey of fast exponentiation algorithms. *Journal of Algorithms*, 27:129–146, 1998.

[GRR98] Rosario Gennaro, Michael O. Rabin, and Tal Rabin. Simplified vss and fact-track multiparty computations with applications to threshold cryptography. In *ACM Symposium on Principles of Distributed Computing*, pages 101–111, 1998.

[Har87] David Harel. Statecharts: A visual formalism for complex systems. *Science of Computer Programming*, 8(3):231–274, June 1987.

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

[HKP05] David Harel, Hillel Kugler, and Amir Pnueli. Synthesis revisited: Generating statechart models from scenario-based requirements. In *Formal Methods in Software and Systems Modeling*, pages 309–324, 2005.

[Hoa85] C. A. R. Hoare. *Communicating Sequential Processes*. Prentice-Hall, 1985.

[Hui04] Huining Feng. DCHARTS, a formalism for modeling and simulation based desgin of reactive software systems. Master's thesis, McGill University, 2004.

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

[ISO89] ISO. Lotos — a formal description technique based on the temporal ordering of observational behaviour. ISO IS 8807, 1989.

[JAP] JAP Anonymity & Privacy. `http://anon.inf.tu-dresden.de/`.

[KAP02] Dogan Kesdogan, Dakshi Agrawal, and Stefan Penz. Limits of anonymity in open environments. In *Information Hiding*, volume LNCS 2578, pages 53–69. Springer-Verlag, 2002.

[KEB98] Dogan Kesdogan, Jan Egner, and Roland Büschkes. Stop-and-go-mixes providing probabilistic anonymity in an open system. In *Information Hiding*, pages 83–98, 1998.

[KNN07] E. Käsper, V. Nikov, and S. Nikova. Strongly multiplicative hierarchical threshold secret sharing. Technical report, submitted to International Conference on Information Theoretic Security (ICITS), 2007.

[Low96] Gavin Lowe. Breaking and fixing the Needham-Schroeder public-key protocol using FDR. In *Tools and Algorithms for the Construction and Analysis of Systems (TACAS)*, volume 1055, pages 147–166. Springer-Verlag, Berlin Germany, 1996.

[LRSW99] Anna Lysyanskaya, Ronald Rivest, Amit Sahai, and Stefan Wolf. Pseudonym systems. In *SAC '99: Proceedings of the 6th Annual International Workshop on Selected Areas in Cryptography*, volume LNCS 1758, pages 184–199. Springer-Verlag, 1999.

[LV07] Mohamed Layouni and Hans Vangheluwe. Anonymous k-show credentials. In *Fourth European PKI Workshop: Theory and Practice*, 2007. *accepted.*

[MH00] U. Maurer M. Hirt. Player simulation and general adversary structures in perfect multiparty computation. *J. of Cryptology*, 13:31–60, 2000.

[MPW89] Robin Milner, Joachim Parrow, and David Walker. A calculus of mobile processes, parts I and II. Reports ECS-LFCS-89-85 and 86, Computer Science Dept., University of Edinburgh, March 1989.

[MSKV06] Sadaf Mustafiz, Ximeng Sun, Jörg Kienzle, and Hans Vangheluwe. Model-driven assessment of use cases for dependable systems. In *MoDELS*, pages 558–573, 2006.

[MSKV07] Sadaf Mustafiz, Ximeng Sun, Jörg Kienzle, and Hans Vangheluwe. Model-driven requirements assessment for dependable systems. *Journal on Software & System Modeling*, 2007.

[Nae06] Vincent Naessens. *A Methodology for Anonymity Control in Electronic Services using Credentials.* PhD thesis, Department of Computer Science, K.U.Leuven, Leuven, Belgium, june 2006.

[NDD06] Vincent Naessens and Bart De Decker. A methodology for designing controlled anonymous applications. In *Security and Privacy in Dynamic Environments*, volume 201/2006 of *IFIP International Federation for Information Processing*, pages 111–122. IFIP TC11, Springer Boston, 2006. , DOI: 10.1007/0-387-33406-8.

[NDDD06a] Vincent Naessens, Liesje Demuynck, and Bart De Decker. A fair anonymous submission and review system. Report CW 442, K.U.Leuven, Department of Computer Science, April 2006. URL: http://www.cs.kuleuven.be/publicaties/rapporten/cw/CW442.abs.html.

[NDDD06b] Vincent Naessens, Liesje Demuynck, and Bart De Decker. A fair anonymous submission and review system. In *Communications and Multimedia Security 2006*, volume 4237 of *Lecture Notes in Computer Science*, pages 43–53. IFIP TC-6 TC-11, Springer, 2006.

[NN05a] V. Nikov and S. Nikova. On a relation between verifiable secret sharing and a class of error-correcting codes. In *LNCS*, volume 3969, pages 275–290. Springer, 2005.

[NN05b] S. Nikova and V. Nikov. On a relation between verifiable secret sharing and a class of error-correcting codes. In S. Nikova, B. Preneel, L. Storme, and J.A. Thas, editors, *Contact Forum*, pages 59–75. Royal Flemish Academy for Science and the Arts, Brussels, 2005.

[OS05] Rafail Ostrovsky and William E. Skeith. Private searching on streaming data. In *In Advances in Cryptology - Proceedings of CRYPTO'05*, volume LNCS 3621, pages 223–240. Springer-Verlag, 2005.

[PC95] European Parliament and European Council. Directive 95/46/ec. *Official Journal of the European Communities*, No L 281:31, 1995.

[PGS04] Ari Juels Philippe Golle, Markus Jakobsson and Paul Syverson. Universal re-encryption for mixnets. In *Topics in Cryptology, RSA'04*, volume LNCS 2964, pages 163–178. Springer-Verlag, 2004.

[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.

[Pla] Platform for privacy preferences (p3p) project. URL: http://www.w3.org/P3P/.

[Pro] KeyLog Pro. `http://www.keylogpro.com/`.

[PTZ03] HweeHwa Pang, Kian-Lee Tan, and Xuan Zhou. Stegfs: A steganographic file system. In *Proceedings of the 19th International Conference on Data Engineering*, pages 657–667. IEEE Computer Society, 2003.

[PV07] Ernesto Posse and Hans Vangheluwe. kiltera: a simulation language for timed, dynamic structure systems. In *Proceedings of the 40th Annual Simulation Symposium*, 2007.

[PvH04] Holger Pfeifer and Friedrich W. von Henke. Formal modelling and analysis of fault tolerance properties in the time-triggered architecture. In *5th Symposium on Formal Methods for Automation and Safety in Railway and Automotive Systems*, 2004.

[Riz97] Luigi Rizzo. Effective erasure codes for reliable computer communication protocols. *Computer Communication Review*, 27(2):24–36, 1997.

[Rr88] G. M. Reed and A. W. roscoe. A timed model for communicating sequential processes. *Theoretical Computer Science*, 58:249–261, 1988.

[RW06] A. W. Roscoe and Zhenzhong Wu. Verifying statemate statecharts using csp and fdr. In Zhiming Liu and Jifeng He, editors, *ICFEM*, volume 4260 of *Lecture Notes in Computer Science*, pages 324–341. Springer, 2006.

[SD02] 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.

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

[Sha79] A. Shamir. How to share a secret. *Commun. ACM*, 22:612–613, 1979.

[SHJ$^+$02] Oleg Sheyner, Joshua Haines, Somesh Jha, Richard Lippmann, and Jeannette M. Wing. Automated generation and analysis of attack graphs. In *SP '02: Proceedings of the 2002 IEEE Symposium on Security and Privacy*, page 273, Washington, DC, USA, 2002. IEEE Computer Society.

[Sim90] Gustavus J. Simmons. How to (really) share a secret. In *Advances in Cryptology - CRYPTO '88*, volume 403 of *Lecture Notes in Computer Science*, pages 390–448. Springer, 1990.

[SMKD05] Aaron Shui, Sadaf Mustafiz, Jörg Kienzle, and Christophe Dony. Exceptional use cases. In *MoDELS*, pages 568–583, 2005.

[Sun07] Ximeng Sun. A Model-Driven Approach to Scenario-Based Requirements Engineering. Master's thesis, McGill University, 2007.

[SV] Ximeng Sun and Hans Vangheluwe. Transforming software requirements by meta-modelling and graph transformation. In *Model Driven Engineering Languages and Systems: 10th International Conference*. MoDELS 2007. Submitted.

[Tas04] Tamir Tassa. Hierarchical threshold secret sharing. In *Theory of Cryptography Conference, TCC 2004*, volume 2951 of *Lecture Notes in Computer Science*, pages 473–490. Springer, 2004.

[TDP07] Carmela Troncoso, Claudia Diaz, and Bart Preneel. Traffic analysis attacks on a continuously-observable steganographic file system. Technical report, COSIC, ESAT, KU Leuven, February 2007.

[UPr] The U-Prove SDK. `http://www.credentica.com/uprove\_sdk.html`. link functional as of Feb 2007.

[VDD06] Kristof Verslype and Bart De Decker. A flexible and open drm framework. In *Communications and Multimedia Security 2006*, volume 4237 of *Lecture Notes in Computer Science*, pages 173–184. IFIP TC-6 TC-11, Springer, 2006.

[VdL03] Hans Vangheluwe and Juan de Lara. Foundations of multi-paradigm modeling and simulation: computer automated multi-paradigm modelling: meta-modelling and graph transformation. In *Winter Simulation Conference*, pages 595–603, 2003.

[VGDD07] Kristof Verslype, Steven Gevers, and Bart De Decker. Towards a privacy-friendly next generation electronic identity card. In *7th Workshop on Privacy Enhancing Technologies*, 2007. *submitted.*

[Whi05] Jon Whittle. Specifying precise use cases with use case charts. In *MoDELS Satellite Events*, pages 290–301, 2005.

[Win05] Windows media drm 10 cracked?, 2005. URL: http://www.engadget.com/2005/02/01/windows-media-drm-10-cracked/.

[YLWD05] Wing Lok Yeung, Karl R. P. H. Leung, Ji Wang, and Wei Dong. Improvements towards formalizing uml state diagrams in csp. In *APSEC*, pages 176–184. IEEE Computer Society, 2005.

[ZMVK05] Miriam Zia, Sadaf Mustafiz, Hans Vangheluwe, and Jörg Kienzle. A modelling and simulation based approach to dependable system design. In Lionel Briand and Clay Williams, editors, *Model Driven Engineering Languages and Systems: 8th International Conference*, volume 3713, pages 217–231. MoDELS 2005, Spring-Verlag, 2005.

[ZMVK07] Miriam Zia, Sadaf Mustafiz, Hans Vangheluwe, and Jörg Kienzle. A modelling and simulation based process for dependable systems design. *Journal on Software & System Modeling*, 2007.

[ZPK76] Bernard P. Zeigler, Herbert Praehofer, and Tag Gon Kim. *Theory of Modeling and Simulation.* Academic Press, first edition, 1976.

[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.

[ZPV] Miriam Zia, Ernesto Posse, and Hans Vangheluwe. Verification of security requirements through multi-formalism modelling and model transformation. In *2nd International Conference on Software and Data Technologies.* ICSOFT 2007. Submitted.