Privacy and Identity Management in Europe for Life

First report on mechanisms

Editors:	Jan Camenisch (IBM)
	Pierangela Samarati (UNIMI)
Reviewers:	Simone Fischer-Hübner (KAU)
	Maren Raguse (ULD)
Identifier:	D2.1.1
Type:	Deliverable
Version:	1.0
Class:	Public
Date:	February 27, 2009

Abstract

Today's society places great demand on the dissemination and sharing of information. Such a great availability of data together with the increase of computational power available today, puts at great risk the privacy of individuals. In fact, privacy is repeatedly identified as one of the main concerns in the global interconnected society. One of the main challenges is then to enable the legitimate use and sharing of information while at the same time guaranteeing both proper protection of the privacy of the individuals to whom information refers and proper preservation of the user's authority over the data.

This document describes the research results obtained by the four work packages of Activity 2, which represent a first step towards the resolution of the issues above. The document includes one chapter for each work package that briefly describes the main research results obained in the first year of PrimeLife along with an indication of what are the issues that will be addressed in the remaining years of the project. The last chapter lists the abstracts of the research papers reporting the findings of the work packages of Activity 2.

The research leading to these results has received funding from the European Community's Seventh Framework Programme (FP7/2007-2013) under grant agreement n° 216483 for the project PrimeLife.

Members of the PrimeLife Consortium

1.	IBM Research GmbH	IBM	Switzerland
2.	Unabhängiges Landeszentrum für Datenschutz	ULD	Germany
3.	Technische Universität Dresden	TUD	Germany
4.	Karlstads Universitet	KAU	Sweden
5.	Università degli Studi di Milano	UNIMI	Italy
6.	Johann Wolfgang Goethe - Universität Frankfurt am Main	GUF	Germany
7.	Stichting Katholieke Universiteit Brabant	TILT	Netherlands
8.	GEIE ERCIM	W3C	France
9.	Katholieke Universiteit Leuven	K.U.Leuven	Belgium
10.	Università degli Studi di Bergamo	UNIBG	Italy
11.	Giesecke & Devrient GmbH	GD	Germany
12.	Center for Usability Research & Engineering	CURE	Austria
13.	Europäisches Microsoft Innovations Center GmbH	EMIC	Germany
14.	SAP AG	SAP	Germany
15.	Brown University	UBR	USA

***Disclaimer**: The information in this document is provided "as is", and no guarantee or warranty is given that the information is fit for any particular purpose. The below referenced consortium members shall have no liability for damages of any kind including without limitation direct, special, indirect, or consequential damages that may result from the use of these materials subject to any liability which is mandatory due to applicable law. Copyright 2008 by IBM Research GmbH, Unabhängiges Landeszentrum für Datenschutz, Technische Universität Dresden, Karlstads Universitet, Università degli Studi di Milano, Johann Wolfgang Goethe - Universität Frankfurt am Main, Stichting Katholieke Universiteit Brabant, GEIE ERCIM, Katholieke Universiteit Leuven, Università degli Studi di Bergamo, Giesecke & Devrient GmbH, Center for Usability Research & Engineering, Europäisches Microsoft Innovations Center GmbH, SAP AG, Brown University.*

List of Contributors

Contributions from several PrimeLife partners are contained in this document. The following list presents the contributors for the chapters of this deliverable.

Chapter	**Author(s)**
Chapter 1: Introduction	IBM, UNIMI
Chapter 2: Cryptographic mechanisms (WP2.1)	**IBM**, GD, K.U.Leuven, UBR
Chapter 3: Mechanisms supporting users' privacy and trust (WP2.2)	**TUD**, KAU, TILT, ULD
Chapter 4: Privacy of data (WP2.3)	**UNIBG**, UNIMI, SAP, TILT
Chapter 5: Access control for the protection of user-generated data (WP2.4)	**UNIMI**, UNIBG, EMIC, SAP
Chapter 6: Conclusions	IBM, UNIMI
Chapter 7: Abstracts of research papers	GD, KAU, K.U.Leuven, IBM, UNIBG, UNIMI, SAP, TUD, UBR, ULD

Contents

List of Figures

List of Tables

Chapter 1

Introduction

The increased power and interconnectivity of computer systems available today provide the ability of storing and processing large amounts of data, often resulting in large collections of data accessible from anywhere at any time. It is also becoming increasingly easier to exchange, access, process, and link such large collections of information. In fact, a user can access voluminous worldwide public information using a standard handheld computer and ubiquitous network resources and, from seeminly innocuous anonymized data and available semi-public information, draw damaging inferences about sensitive information on specific individuals and/or organizations. For instance, data holders often release information with all explicit identifiers (e.g., name, SSN, and address) removed, incorrectly believing to safeguard the anonymity of the parties to whom the released information refer. However, released information often contains other data (e.g., birth date and ZIP code) that can be linked to publicly available information to re-identify individuals. The solution to this problem cannot be to entirely avoid data release also because the Society has developed an interest for all kinds of detailed information for many purposes, and modern systems tend to distribute information widely. Furthermore, when releasing some information: *1)* we often lose control of how this information is used, and how, and to whom it is disclosed; and *2)* while before we needed to trust a specific person or organization, we now have to worry about putting trust, or some control, over the entire global interconnected society. It is therefore inevitable that we have an increasing degree of awareness with respect to privacy. Privacy issues have then been the subject of public debates and discussions and many controversial proposals for the use of information have been debated openly.

Since privacy by its own nature is a multifacet concept that may encompass several meanings, depending on different contexts, PrimeLife's main focus is not only on the technological aspect of privacy within today's global interconnected society, where users interact with remote information sources for retrieving data or for using on-line services, but also on the legislation (law and public policy), organizational, and on individual policies and practices. This raises a number of research challenges, some of which will be addressed in Activity 2, whose general goal is the development and application of new privacy-aware techniques addressing the problem of guaranteeing privacy and trust in the electronic society. More precisely, the research challenges addressed in Activity 2 can be summarized as follows.

- *Providing and supporting life-long privacy.* There are many emerging scenarios (e.g., collaborative environments) where users reveal personal information that is then used for profiling to provide services that are better suited to the need of the users. This situation however introduces new privacy threats since users can be tracked and their activities can be monitored. To avoid these issues, cryptographic schemes and protocols (e.g., anonymous credentials, delegation of credentials, searchable encryption) that support users in protecting their personal privacy through data release minimization and that can provide necessary trust guarantees are becoming even more important.

- *Supporting users' control over their data.* One of the most important privacy protection principles states that personal information collected for one purpose may not be used for any other purpose without the specific consent of the person it concerns or a legal provision allowing to process the data for a different purpose. However, users often provide personal information for use in one specific context having no idea how such personal information may be used subsequently, thus losing the control over it. It is then important to develop novel mechanisms and tools that allow users to verify whether their information is managed according to privacy law regulations and privacy constraints specified by the users themselves.

- *Assessing and efficiently enforcing privacy.* Existing approaches for protecting personal data often do not provide any mechanism for measuring the protection provided in a release and the precision and usefulness of the resulting data. As a matter of fact, protection can make the data not more useful for the purpose they were required. Determining "optimal disclosure results" in terms of privacy and utility requires then new insight into measuring the usefulness of released data and the effectiveness of the protection provided. It is then important to introduce effective measures for privacy that can be used by data holders and users to assess the protection of data from improper exposure as well as privacy protection techniques able to meet the desidered level of privacy.

- *Supporting the protection of user-generated data.* Large collections of personal data are often generated from and/or by users. Existing approaches for protecting such collections of personal data often are based on the assumption that the data holder (i.e., the party storing and managing access to the data) is a trusted entity, leaving to the data holder the realization of access restrictions. There are however many emerging scenarios (e.g., outsourcing scenarios) where such an assumption does not hold anymore. It is then important to develop novel techniques that allow users to effectively gain control and that allow the dissemination of data by a data holder that may offer a data storage service without having the permission of accessing the data.

Activity 2 is organized in four work packages that address the above research challenges. *Work Package 2.1: Cryptographic mechanisms* focuses on cryptographic techniques for supporting privacy and trust. *Work Package 2.2: Mechanisms supporting* *users' privacy and trust* focuses on solutions for supporting users in checking whether their personal data are used in accordance withprivacy laws and privacy constraints specified by them. *Work Package 2.3: Privacy of data* focuses on solutions for assessing and ensuring privacy of large collections of sensitive data. *Work Package 2.4: Access control for the protection of user-generated data* focuses on solutions for enabling the enforcement of access restrictions on data generated from and/or by users.

The research results achieved in these work packages during the first year of PrimeLife have been presented in international conferences, including leading conferences such as ACM CCS, ICDCS, and ASIACRYPT. The main research results of each work package are summarized in the following.

- *WP2.1: Cryptographic mechanisms.* This work package has produced results in seven different areas: *1)* anonymous credential systems; *2)* protocols for accessing services without revealing the users' access patters; *3)* selective access control in social networks; *4)* privacy-aware third party services; *5)* protocols supporting a private intersection of certified sets; *6)* biometric algorithms for embedded systems; *7)* high level security compartments of identity management wallets.

- *WP 2.2: Mechanisms supporting users' privacy and trust.* This work package has produced results in five different areas: *1)* linkage control and mechanisms supporting transparency; *2)* privacy measurements for assessing the privacy of a user in communication systems; *3)* privacy-aware collaborative groups; *4)* privacy-aware interoperable reputation systems; *5)* mechanisms for supporting privacy awareness.

- *WP2.3: Privacy of data.* This work package has produced results in five different areas: *1)* privacy assessment; *2)* privacy in mobile environments; *3)* privacy metrics and privacy enforcing methods (anonymization) to assess the privacy protection enjoyed by data collections; *4)* privacy in business applications; *5)* specification and enforcement of privacy constraints in relational databases.

- *WP2.4: Access control for the protection of user-generated data.* This work package has produced results in five different areas: *1)* protection of location information; *2)* secondary use restrictions; *3)* confidentiality of privacy policies; *4)* management and protection of biometric traits; *5)* dynamic access control mechanisms.

The remainder of this document is organized in four chapters, one for each work package (Chapters 2-5). Each chapter describes the research work performed in the first year of PrimeLife as well as work planned for subsequent years. Chapter 7 reports the abstracts of the research papers reporting the findings of the work packages throughout the reporting period.

Chapter 2

Cryptographic mechanisms (WP2.1)

2.1 Introduction

Cryptographic schemes and protocols that support users in protecting their personal privacy gain in importance because most of our personal data are nowadays stored and processed in digital format. Encryption, hashing, and simple authentication techniques form the basis for secure information processing systems. However more advanced cryptographic mechanisms are needed if we desire security properties that go beyond the simple confidentiality and authenticity properties. This holds particularly true, if no single party is trusted by all the users, and all users want to retain some control over their data. Such protocols include, but are not limited to, anonymous credentials [Cha85, CL01a, Lys02], electronic cash [Cha82, Bra93, CHL05], group signatures [CvH91, CS97, ACJT00, BBS04], private information retrieval, oblivious transfer, blind signatures [Cha82], and, in the most generalized case, secure multi-party computation of any efficiently computable function.

Work Package 2.1 is concerned with research that distills and if necessary extends such mechanisms into a toolkit for solving privacy-enhancing identity management problems.

In today's electronic transactions, delegation of credentials is increasingly necessary; without delegation, only the most stringent access control policies can be realised; those that are actually in use today do require delegation. Delegation captures trust relationships on which communities are built. Lack of efficient cryptographic algorithms for delegation of anonymous credentials is an impediment for adopting anonymous credentials in practice, and stands in the way of making privacy-preserving protocols the default.

Furthermore, the scenarios of PrimeLife's Activity 1 raise a number of open cryptographic problems for which no solutions exist. PrimeLife is concerned with issues such as social networks privacy, life-long privacy, and the deployment of privacy mechanisms in the real world through the means of open source initiatives and standardization. Mechanisms such as delegation of credentials, searchable encryption, and oblivious service access, prove useful within the thick social fabric of today's Internet where private information needs not only be protected but, in addition, needs to be processed and shared in a responsible manner.

Examples for such mechanisms are signature schemes that allow one to copy only parts of a signed text such that the recipient can still be convinced that the part originated from an anonymous but trustworthy source. Another example are special encryption and key management schemes that allow users to enforce, in a distributed fashion (i.e., without any trusted components such as a central server or DRM-like hardware), that their data can only be read by specific users, e.g., their friends.

Work of this work package is divided into two tasks. The first one, *Task 2.1.1: Cryptography for Privacy and Trust*, is concerned with the study and design of cryptographic algorithms for privacy and trust. This also includes algorithms for biometric authentication. The second task, *Task 2.1.2: Trusted Wallet* is focused on the design and development of components that allow users to securely manage their cryptographic key materials such as secret keys and electronic credentials.

2.2 Research results

2.2.1 Cryptography for privacy and trust (Task 2.1.1)

In the reporting period, Work Package 2.1 has improved the state of the art in the following major areas.

Anonymous credentials. We have continued our research on new credential systems to improve both the efficiency of the attribute encoding and the revocation of credentials.

Private service access. In certain sensitive cases users may be interested in using a service unobservably, i.e., even the service provider would not learn the exact type of service requested or the query that is answered. This is important to protect against insider attacks, especially if the service offered is a function of personal information of other data subjects, e.g., as it is the case for social network sites. We investigated ways to incorporate privacy friendly service selection and payment protocols into such a system. In particular we provide protocols for searching on encrypted data and priced oblivious transfer. From a more theoretical angle we are looking for oblivious transfer protocols that fulfill the strongest notions of security for the composition of multiple protocols.

Delegation. While the delegation of non-anonymous credentials is easily solved and widely used in the form of a certification hierarchy, a solution to this problem for anonymous credentials proved to be evasive. We obtained a feasibility result using the Groth and Sahai proofs [GS08].

Cryptography for selective access control in social networks. The aim of cryptographic support for selective access control in social networks is to give the user control not only about the definition of the access control policy, but also its enforcement. In a first step, we will develop policy concepts that are general enough to describe access control restrictions for a variety of different social networking sites. We will also investigate different means for enforcing such policies using sticky policies and advanced encryption techniques.

Privacy aware services. Offering services in a privacy friendly way is often very challenging and requires new cryptographic mechanisms and protocols. The goal here is to achieve privacy while maintaining accountability using new protocols but also potentially by relying on third parties to enforce security properties. Examples of such parties could be identity escrow services. Thereby, these third parties should not be involved in the normal operation of the system but only in case some party misbehaves. Also, the trust that other parties need to put into such third party should be minimal. Ideally, the third party should be completely accountable, i.e., everyone should be able to verify that the party behaved as specified without being able to derive any other information. The work package studies different ways to offer services and how third parties can best be employed and first results here have been published.

Biometric algorithms. Biometric user authentication is becoming popular in various applications. Most scenarios are designed to store biometric credentials such as fingerprints in a system global database. It must be understood that biometric data are extremely sensitive to identity fraud and identity theft. Biometric data are not secrets but unique identifiers. Once compromised, there is no possibility to restore or regenerate since the credentials are characteristic for the individual – usually over a lifetime. The goal is to enhance privacy enabling technologies such as Match-on-Card and make these more convenient and practical to use. The storage and comparison within a portable embedded system makes biometric data less vulnerable against attacks and allows end users better control in managing sensitive private data.

The results of our work are described under separate subsections below.

Anonymous credentials

A *credential system* is a system in which users can obtain credentials from organizations and demonstrate possession of these credentials. For instance, such a credential could be a driver's license containing as attributes data about the user such as her birthday, address, or the date she took the driving test. Other examples include passports, identity cards, or educational certificates. That is, a credential is a signed statement issued by some party (issuer) about another party (recipient or user).

Unlike a traditional PKI, in an *anonymous credential system*, such statements can be presented to third parties in a way such that the issuing and the presenting transaction are not linkable. Moreover, if the user wants to use her driver's licence to show that she is of age, a credential system allows her to do so without revealing any other information about her. In other words, such systems allow the user to enforce what information another party learns about her. Thus, such credential systems form the foundation of any privacy enhancing identity management system and hence one of the goals of the work package is to improve the state of the art in this space. This includes to make the existing protocols more efficient, realize new features as to enable new application, and to provide schemes that are based on simpler cryptographic assumptions.

In the first project year, we have achieved two main break through results in the area of anonymous credential systems: (1) efficiency for credentials with a large number of attributes (with a relatively small domain), (2) efficient revocation with fast witness updates.

Efficient Attribute Encodings. We extend the Camenisch-Lysyanskaya anonymous credential system such that selective disclosure of attributes becomes highly efficient. The resulting system significantly improves upon existing approaches, which suffer from a linear complexity in the total number of attributes. This limitation makes them unfit for many practical applications, such as electronic identity cards. Our system can incorporate an arbitrary number of binary and finite-set attributes without significant performance impact. Our approach folds all such attributes in a single attribute base and, thus, boosts the efficiency of *all* proofs of possession. The core idea is to encode discrete binary and finite-set attribute values as prime numbers. We use the divisibility property for efficient proofs of their presence or absence. We additionally contribute efficient methods for conjunctions and disjunctions.

Efficient Range Proofs. An often met requirement for anonymous credentials is that a user can prove that a value contained in her credential lies in a given range or is a member of a public set. For instance, to access a teenage chat room, a user might need to prove that her eID card contains a birthdate showing that she is between 12 and 15 years old. So, in general, we require an efficient proof that certified value is a member of some public set Φ. We have presented two new approaches to building set-membership proofs. The first is based on bilinear group assumptions. When applied to the case where the set Φ is a range of integers, our protocols require $O(\frac{k}{\log k - \log\log k})$ group elements to be exchanged. Not only is this result asymptotically better, but the constants are small enough to provide significant improvements even for small ranges. Indeed, for a discrete logarithm based setting, our new protocol is an order of magnitude more efficient than previously known ones. We have also presented alternative implementations of our membership proof based on the strong RSA assumption. Depending on the application, e.g., when Φ is a published set of values such as frequent flyer clubs, cities, or other ad hoc collections, these alternatives also outperform prior solutions.

Efficient Revocation. The success of electronic authentication systems, be it e-ID card systems or Internet authentication systems such as CardSpace, highly depends on the provided level of user privacy. Thereby, an important requirement is an efficient means for revocation of the authentication credentials. In this paper we consider the problem of revocation for certificate-based privacy-protecting authentication systems. To date, the most efficient solutions for revocation for such systems are based on cryptographic accumulators. Here, an accumulate of all currently valid certificates is published regularly and each user holds a *witness* enabling her to prove the validity of her (anonymous) credential while retaining anonymity. Unfortunately, the users' witnesses must be updated at least each time a credential is revoked. For the know solutions, these updates are computationally very expensive for users and/or certificate issuers which is very problematic as revocation is a frequent event as practice shows.

We have proposed a new dynamic accumulator scheme based on bilinear maps and shown how to apply it to the problem of revocation of anonymous credentials. In the resulting scheme, proving a credential's validity and updating witnesses both come at (virtually) no cost for credential owners and verifiers. In particular, updating a witness requires the issuer to do only one multiplication per addition or revocation of a credential and can also be delegated to untrusted entities from which a user could just retrieve the updated witness. We believe that thereby we provided the first authentication system offering privacy protection suitable for implementation with electronic tokens such as eID cards or driver's licenses.

Private service access

We are interested in protocols that allow users to access services without revealing their access patterns. Such techniques are known as oblivious transfer and Private Information Retrieval (PIR).

Oblivious access to protected resources could for instance be used to realize social networking services in which the underlying social network is itself hidden from the service provider while users can still find out information about the friends of their friends. Another scenario is sensitive databases, e.g., DNA databases. A user that contributed her DNA to a DNA database has a right to be notified about every use of her data. However the fact that a medical institute accessed a certain piece of DNA may already be extremely sensitive and may need to be hidden even from the administrator of the database. The issues get more contrived as service providers need to control access to the database, e.g. to protect the privacy of the data subject or to selectively charge money for the access to certain data.

In the first project year, we investigate the case in which we assign each message of an Oblivious Transfer (OT) a price that the user has to pay. Priced Oblivious Transfer (POT) is a generalization of OT where the receiver pays for the messages without the sender learning the amount paid. Both OT and POT admit an adaptive variant where the receiver chooses σ_i after receiving $m_{\sigma_{i-1}}$, which enables applications such as privacy-preserving e-commerce.

We present an adaptive OT scheme and we further modify it to construct an adaptive POT scheme. Both constructions are universally composable, optimal in terms of rounds of communication and, after an initialization phase of complexity $O(N)$, both have constant communication and computational cost in each transfer phase. This work has not been published yet.

Another problem related to oblivious transfer is oblivious searching. Instead of receiving the message corresponding to a hidden index, the user receives all messages that match with a hidden search keyword. We show how oblivious keyword search can be implemented for public key encrypted data [CKRS09].

Delegation

A recurring problem in identity management systems is the delegation of access rights. Delegation is a common tool in the physical world when one lends to someone else one's keys or when one signs a paper document that delegates certain rights (for example to vote, to buy a house, or to submit a proposal). In the on-line world this can be achieved by passing on a password, but it is clear that this brings unintended consequences. Moreover, there are many complex ways of delegating access rights or credentials: there are delegations with and without restrictions in scope, with and without anonymity, delegations that can be passed on themselves etc.

Conventional public key certificates support delegation. A certification authority can certify a company, and the company can in turn certify its employees. These certificates do however leak the identity of all intermediary certificates.

Traditional anonymous credentials do not support delegation, and the implementation of delegation properties similar to those described above are difficult to achieve. We provide some first results to allow for anonymous credentials that allow to prove statements such as the following: "My company has a subscription to your database, and has granted me access to a specific subset of the database", without revealing the name of my company.

In the first project year, we revise the entire approach to constructing anonymous credentials and identify *randomizable* zero-knowledge proof of knowledge systems as the key building block [BCC$^+$08]. We formally define the notion of randomizable non-interactive zero-knowledge proofs, and give the first construction by showing how to appropriately rerandomize Groth and Sahai (Eurocrypt 2008) proofs. Our insight is that instead of giving Alice his signature, Oliver gives Alice a non-interactive proof-of-knowledge of the signature. The trick is to find a proof-system that would then let Alice (1) delegate the credential by extending the proof and (2) rerandomize the proof every time she shows (or extends) it to preserve her anonymity.

Let's say Oliver is a credential authority with public key pk_O and secret key sk_O; and let's say Alice is a user with secret key sk_A. Alice wants to obtain the credential directly from Oliver (so her certification chain will be of length 1). Under the old approach, they would run a secure two-party protocol as a result of which Alice obtains a signature $\sigma_{pk_O}(sk_A)$ on sk_A, while Oliver gets no output. Under the new approach, Alice's output is $(comm_A, \pi_A)$, where $comm_A$ is a commitment to her secret key sk_A, and π_A is a *proof of knowledge* of Oliver authenticating the contents of $comm_A$. Note that a *symmetric* authentication scheme is sufficient because *no one ever sees the authenticator*; all verification is done on the proof of knowledge. The symmetric key sk_O remains secret to Oliver; we create a "public" key C_O that is simply a commitment to sk_O.

How can Alice use this credential anonymously? If the underlying proof system is *malleable* in just the right way, then given $(comm_A, \pi_A)$ and the opening to $comm_A$, Alice can compute $(comm'_A, \pi'_A)$ such that $comm'_A$ is another commitment to her sk_A that she can successfully open, while π'_A is a proof of knowledge of Oliver authenticating the contents of $comm'_A$. Malleability is usually considered a bug rather than a feature. But in combination with the correct extraction properties, we still manage to guarantee that these randomizable proofs give us a useful building block for the construction. The bottom line is that $(comm'_A, \pi'_A)$ should not be linkable to $(comm_A, \pi_A)$, and also it should not be possible to obtain such a tuple without Oliver's assistance.

How does Alice delegate her credential to Bob? Alice and Bob can run a secure protocol as a result of which Bob obtains $(comm_B, \pi_B)$ where $comm_B$ is a commitment to Bob's secret key sk_B and π_B is a proof of knowledge of an authenticator issued by the owner of $comm'_A$ on the contents of $comm_B$. Now, essentially, the set of values $(comm'_A, comm_B, \pi'_A, \pi_B)$ together indicate that the owner of $comm'_A$ got a credential from Oliver and delegated to the owner of $comm_B$, and so it constitutes a proof of possession of a certification chain. Moreover, it hides the identity of the delegator Alice! Now Bob can, in turn, use the randomization properties of the underlying proof system to randomize this set of values so that it becomes unlinkable to his original pseudonym $comm_B$; he can also, in turn, delegate the credential to Carol.

Cryptography for selective access control in social networks

Anonymous credentials are not a perfect fit for the privacy needs of social network (SN) sites. The users of social networks do not aim at data minimization. They are using it to share information. As we will see there will still be some legitimate uses of anonymous credentials in social networks but a different core strategy is needed.

The main privacy concern of SN users is not data minimization but control. Users want to specify who can see which information (specify access control policies) and have them enforced by the SN, preferably without putting too much trust in external parties (everyone who is not them, or their peers).

The main mechanism to enforce access control is a reference monitor. The main mechanism to reduce trust in external parties is to (i) either store data locally (or at a proxy that they trust) which would require users to be always online (or invest in additional infrastructure), or (ii) store data in encrypted form. We consider the latter approach. The reference monitor and the encryption need to be integrated with each other in an adequate way. We consider the following option.

- Proxy reencryption: The social network server cannot decrypt himself, but can translate between encryptions for different users.
- Broadcast encryption: A message can be encrypted to a large number of users without increasing the cipher text size.
- Ciphertext-Policy attribute-based encryption (CP-ABE): CP-ABE is a variant of Broadcast encryption in which the set of users that can decrypt a message is described by a predicate over the attributes of the user. Users that have adequate attributes can decrypt, others cannot. This is a natural way for implementing sticky policies. The attributes “friend”, “close friend”, “family” might be useful attributes, or “male” and “female”. In order to encrypt to all her female friends Alice could use the policy “friend and female”.
- Broadcast/CP-ABE reencryption: One may want a server to change the set of users that can decrypt a ciphertext without being able to decrypt it herself.
- Key management: Users need to obtain long term secret keys and short term data encryption keys. The second type of keys (short term) would be encrypted using

the schemes sketched above using the long term key. Long term keys can be identity based or PKI based.

We are currently collaborating with Activity 1 to extract additional requirements for such advanced cryptographic techniques for policy enforcement. However, one first needs to exploit the full power of classical hybrid encryption techniques, before it becomes reasonable to start new research in this area.

Privacy aware third-party services

Services with identity management relations have a challenging set of requirements. They require accountability of identity data as well as privacy protection of their users. The first requirement is usually driven by the need to mitigate risk to the service provider, and the latter requirement driven by regulatory compliance (e.g., privacy legislation). This requires new primitives for accountable and privacy-preserving identity management methods for services that provide improved support for these two requirements.

Anonymous credential systems exist that provide methods to achieve a wide range of accountability and privacy goals for services. In particular anonymous credentials systems already have the capability to selectively disclose statements about a user, e.g., proving the user's age without revealing the user's actual date of birth. With the complementary primitive of verifiable encryption [CS03], credential systems can also provide accountability without infringing on privacy requirements. For instance, a user U can encrypt her true identity to the authorities, provide this encrypted data to a service provider SP, and convince SP in a zero-knowledge proof of knowledge that this encrypted data contains a valid user identity that can be opened by the authorities.

Existing systems realize accountability with verifiable encryption by encrypting the user's identity to a trusted third party (TTP), for instance, law enforcement authorities. This presents three challenges:

1. This mode of operation involves a fully trusted TTP with no graceful degradation of privacy and security should the TTP become compromised.
2. A malicious service provider holding a verifiable encryption may attempt to betray the user by opening a decryption case at the TTP without good cause.
3. Honest service providers find the traditional system encumbering because of the need to involve such highly trusted authorities for even minor dispute cases. For example, to bring a case to law enforcement in the real world is likely to have a non-trivial cost, both in the time required, and in support from legal council.

Therefore, we propose to rethink the corresponding primitives to better meet the actual needs of services.

In the first project year, we have proposed an improvement of such a system's flexibility and trust model. We propose what we believe to be the first verifiable encryption system to provide revocation of decryptability with end-to-end unlinkability. This simultaneously provides stronger privacy for the user, and a lower cost to the service provider for obtaining accountability when the user misbehaves. In particular, our solution achieves the following properties.

1. Our revocation authority RA is a weaker TTP that cannot link user's transactions within the system. If the RA becomes compromised it is still restricted in what information it can reveal.

    - RA cannot learn any information about the user before the user's bill is past due.
    - RA processes only blinded information. When an anonymity revocation is requested by a service provider, RA only knows that it is checking the key for a legitimate transaction, without knowing which transaction or which user. Therefore, RA cannot block requests selectively or collude against any specific user.
    - Even when the bill is past due, and the user has not paid it, RA cannot link this fact to any particular contract or user.

2. Our system contains a mechanism for verifiable, yet privacy supporting, proof of fulfillment of the contracted terms of the service. The RA can easily detect an unfounded request for opening an encrypted identity, and will not service such requests.

3. Our system permits automatic identity revocation in the event that a contract is not fulfilled by the user. Because the contract satisfaction condition can be machine verified, external authorities such as law enforcement do not need to be involved.

Private intersection of certified sets

The problem of private set intersection is the following. Alice and Bob hold sets S_A and S_B, respectively. They would like to jointly compute the intersection, in such a way that reveals as little as possible about S_A to Bob and S_B to Alice. In other words, both Alice and Bob should learn $S_A \cap S_B$ but nothing more.

Private set intersection protocols may find applications in online recommendation services, medical databases, and many data related operations between companies, which may even be competitors. An example from the law enforcement field is given by Kissner and Song [KS04]; suppose a law enforcement official has a list of suspects and would like to know if any of them are customers of a particular business. To protect the privacy of the other customers, and keep the list of suspects private, the business and the enforcement authority use a private set intersection protocol to learn only those names appearing on both lists.

While the task of computing the intersection could be completed with general secure multiparty techniques, it is far more efficient to have a dedicated protocol, especially since the number of communication rounds will be constant. A number of such protocols exist in the literature. A problem common to all previous protocols is that the inputs S_A and S_B can be chosen arbitrarily by Alice and Bob.

Suppose Bob is malicious in the following sense; he follows the protocol, but wishes to learn as much about S_A as possible. Bob's strategy is to populate a set S'_B with all of his best guesses for S_A and to have $|S'_B|$ be as large as Alice will allow. This maximizes the amount of information Bob learns about S_A.

In the extreme case, Bob may claim S_B contains all possible elements, which will always reveal S_A. He may also vary his set over multiple runs of the protocol, in order to learn more information over time. These attacks are even more powerful when the protocol can be executed anonymously. Note that all this behaviour is permitted in any model which allows the participants to choose their inputs arbitrarily. This weakness of models which allow arbitrary inputs reduces the practicality of private set operations.

In the first project year, we have proposed to use certified inputs to address the problem discussed above and have presented protocols that realize this. The goal of certifying the private sets of participants is to restrict their inputs to “sensible” or “appropriate” inputs. A certification authority (CA) is a trusted party who certifies that each participant’s set is valid. Once the sets are certified, the CA need not be online.

For example, suppose companies want to perform set operations on their financial data. Each company uses a different, but trusted, accounting firm who certifies the data. The companies can then perform as many operations with as many other companies with their certified data.

Since our approach to certifying sets shares a lot with anonymous credentials, this area may also benefit from our work. Credential holders may treat values in their certificates as sets, and intersect them. For example, two pseudonymous/anonymous users may intersect their credentials to determine they live in the same city and were born in the same year. As another example, they may determine whether their ages are within y years by intersecting sets of integers $\{age - y, \ldots, age, \ldots, age + y\}$, where age is the certified value from the credential.

Algorithms for biometry

The major objective in this work package is to enhance biometric algorithms for embedded systems. The storage and verification of biometric credentials should be performed in a tamper-proof embedded system. In a networked world, where users would not trust the communication partner, they still can trust the smart card in their wallet. A number of implementations for on-card biometric algorithms are available today. Fingerprint verification was the first biometric technology to be adopted by the smart card experts [Mue01]. Most on-card algorithms today only focus on comparison of extracted features. The image processing and generation of a characteristic dataset still is done in the host system or backbone. Our work focuses on implementing image processing in smart cards. This enhances the privacy and data management. A fair matching accuracy shall be achieved while keeping the storage requirements and computational complexity in a range suitable for smart cards. We will focus on fingerprint comparison in this work. It is not only the most common biometric trait but also has the largest available databases, best understanding in terms of technology and is one of the most critical to be compromised.

In the first project year, we focused on two key aspects:

- analyzing existing on-card algorithms and understanding the need for a full image processing on-card, and
- real world scenario requirements to realize a stable algorithm under varying circumstances.

The previous work on fingerprint match-on-card included already some proposals or prototypes that go beyond the normal minutia matcher [MM06]. It turned out that the images generated from normal fingerprint sensors are too large to be used in smart cards. The fingerprint image can be reduced in both resolution and color depth without significant loss in quality. While today's sensors operate with 500dpi and 8bit grey levels, the on-card image processing will have to work with 360dpi and 4bit grey levels. The environmental conditions are also a major research aspect. If a system is designed to work fine in the lab, it might still be useless when exposed to a different population or various environmental influences summarized in [FSSRAMM08].

2.2.2 Trusted wallet (Task 2.1.2)

An identity management wallet has a similar function as a wallet in the real world. Instead of authenticating towards the different service providers the user would authenticate to her "trusted identity management wallet" which contains all of her access credentials. By concentrating the sensitive data within a single application the potential risk of harm is increased. We identified three compartments of an identity management system that should be protected independently:

1. The compartment running the applications which eventually want to make use of the IdM system.

2. The compartment doing the processing of the personal information (i.e. the wallet) and the interactions with the user.

3. The compartment containing the high security cryptographic material.

It is preferable to run the different compartments of a trusted wallet separated from each other and with different requirements on their security:

- *Low security.* Application software such as the user's browser should run with normal operation system security. This guarantees good support of legacy software and good integration into the user's familiar computing environment.

- *Medium security.* The wallet software including the identity selector is run in a compartment provided by a separate virtual machine.

- *High security.* A smart card or a threshold scheme that remains secure as long as an adversary does not control a majority of a user's portable devices can be used to secure important cryptographic material.

In the first project year, we focused on the high security compartment. The role of the high level security compartment is to act as a last trust anchor for the user's identity management system. If the user's computer gets hacked, if she loses his laptop, wrote all his passwords on a sheet of paper, or shared his whole wallet with an untrusted person or all her peers in a global file sharing network, the high security compartment will try its best to still provide some protection of a user's credentials and personal data.

This protection will not be perfect, a user can always decide to publish any information about himself over unprotected channels, and he can decide to act as an online proxy that shows credentials for the whole world to allow everyone to impersonate him.

However, the security compartment can still do its best to protect users who do not decide to shoot themselves into the foot. This indicates a second challenge, a tradeoff between security and usability. Adding security requires additional user interaction, e.g., to ask the user whether she wants to release some data, to enter a PIN and so on. If the burden for the user becomes too high, she is likely to stop using the system.

One option for implementing a high security compartment is to make use of distributed cryptography. Typically a user possesses more than one computing device which would allow the distribution of the credential store onto those devices. The goals being that a number of compromised devices do not allow the reconstruction of the credentials as long as the number of compromised devices is below a certain threshold. Reversely, the reconstruction is possible in case a small number of devices has been lost. This restore process is necessary to encourage the use of mobile devices. Consequently, the distributed storage of credentials adds redundancy which is beneficial in case of a device failure (e.g. loss, break down and so fourth). The combination of the information gathered from the different devices is done at the wallet which allows the storage of the credentials to be untrusted.

Another option is to use tamper resistant devices to store the credential master secret. If the tamper resistant device has basic input/output capabilities this allows for additional security measures. The user would be enabled to monitor (i) if the tamper resistant device is taking part in the transaction, (ii) what information is being requested by the trusted wallet and (iii) what personal information is released.

2.3 Future research

2.3.1 Cryptography for privacy and trust (Task 2.1.1)

Anonymous credentials

There are a large number of issues to solve to make anonymous credential more versatile and applicable to a wider range of applications. Issues include the following:

- *Alternative variants for efficient revocation.* Depending on the actual scenarios, the requirements for revocation differ much more than one would expect. We will continue to study different scenarios and their requirements and will try to come up with revocation mechanisms that are targeted to these scenarios.

- *Lightweight devices.* We are going to study how to best achieve the functionality of anonymous credentials for lightweight or restricted devices such as smart cards as using the standard anonymous credential protocols (at least as soon as more than just the very basic functionality is needed) seems just a bit off from what would be accepted by end users in terms of computation times.

Private service access

We have just started the work in this area in the first project year and will continue our work as outlined earlier. That is, we will investigate on how to use anonymous credentials certifying the attributes, roles, and/or access rights of users to authorize access to different resources such that: 1) the service provider will not learn which resource a user accesses and 2) it is at the same time ensured that only authorized users can access a particular resource.

Privacy aware third-party services

Also in this area we have so far only achieved initial results and hence plan to further investigate different scenarios and come up with new or better solutions. The end goal here is to derive a set of standard mechanisms that will allow one to implement any such services in a way that the third party is, one the one hand, accountable, i.e., all other parties can verify that it has performed its task as expected and, on the other hand, is not aware of the identities of the parties involved in the transactions. The latter will ensure that the third party will not be able to change its behavior depending on the specific user involved.

Biometric authentication

The research on both initial aspects will be continued. A first version of embedded image processing and feature extraction software will be implemented in the second year. It is designed to run on a microcontroller with the ARM7 architecture that is available in today's state-of-the-art smart card chips. The optimization will use traditional methods such as a time-memory trade-off. Predefined tables eliminate the need for trigonometry operations that are computationally costly. Parameter tuning will become a key aspect when sampling down raw images. It is not yet foreseeable how to overcome the memory constraints and matching accuracy simultaneously. A scientific paper will be submitted in 2010 discussing the potential and proposed solutions. A physical prototype is planned and will serve as a proof of concept and technology base for products in the long term.

2.3.2 Trusted wallet (Task 2.1.2)

In this area we have just started the work in the first project year and will continue as outlined earlier.

In particular, we will focus on two of the described compartments as the component having usual user level privileges is not to be substantially changed by the wallet. Due to their different capabilities compartments with more restrictions are encapsulated by the less protected domains. This difficulty has to be overcome to attain guarantees about the correct execution and protection of the highly sensitive data. More specifically, the medium security compartment requires a confidential and authentic channel to the service provider. Similarly, the high security compartment can only work properly and in the user's interest if the integrity of the medium security layer is guaranteed. This is particularly relevant if this component does not have dedicated I/O possibilities.

We will continue exploring the trade-off between security and usability. It might be possible to get a deeper knowledge of the consequences of people's actions which could counter the negative effects that the additional security has on the usability. For example, allowing a user to take an informed decision by highlighting the downsides before accessing a service might be considered as an improvement of the state-of-the-art. Capabilities of the wallet in this area arise due to the comprehensive approach of consolidating all the user's accounts.

The prototype will focus on the mentioned security aspects and implement one possible solution. It will show how the most critical requirements can be met along with how a trusted wallet can be employed to accomplish a smooth user experience.

Chapter 3

Mechanisms supporting users' privacy and trust (WP 2.2)

3.1 Introduction

The Internet offers its users numerous possibilities to interact with each other. Interactions cover various fields of interest and parts of life for many people. Examples most of us are familiar with are e-shopping, e-health, on line community services and e-government. Protecting privacy is an important issue in electronic interactions. Revealing personal data and being completely unaware of privacy may cause a lot of privacy issues later. The secondary use of data contributes to these problems [Var96]. Within this section we first give an overview on different approaches, which help to preserve or even control the users' privacy. They can be used either in addition or as a part of a privacy-enhancing identity management system (PE-IMS).

To achieve this type of control transparency tools play an important role. Transparency is a legal privacy principle, which also can be derived from the EU Data Protection Directive 95/46/EC [Dir]. When a data controller is requesting personal data from a data subject, the data controller must inform the data subject about her identity (name and address), the purposes of data processing, the recipients of the data and all other information required to ensure the processing is fair ([Dir] Art. 10), The data subject has the right to access all data processed about her, to demand the rectification, deletion or blocking of data that is incorrect or is not being processed in compliance with the data protection rules ([Dir] Art. 12). The user's right to access also includes the right to obtain knowledge of the logic involved in any automatic processing of data concerning her. Even though there is no legal requirement that users can exercise their rights online, we believe that such a state of affairs would be beneficial for all parts involved and could also make the process more administratively efficient. In Section 3.2.1 we present our research results on *transparency support tools*.

Taking into account the definition of privacy as actively experienced *"right to select what personal information about me is known to what people"* [Wes67], privacy awareness encompasses a users' perception, cognition and attention on whether others receive or have received personal information about her, her presence and activities,which personal information others receive or have received in detail, who receives or has received personal information, and how these pieces of information are or might be processed and used. In Section 3.2.5 we describe how to design tools to support *privacy awareness* and also study the influence of privacy awareness support in users' behavior.

Transparency and awareness need a measurement of a user's privacy, especially her level of anonymity and the unlinkability of her usage of services. There have been made several proposals to formalize and measure anonymity and unlinkability for both communication systems and applications. In Section 3.2.2 we give an overview on common approaches on *privacy measurement*.

Interaction systems usually both create and implement a virtual community [Rhe93]. Sharing personal information with others is the basic idea of many social applications like communities to create trust among users. Privacy enhancements tend to hinder trust, because trust usually is built up on information about and from the interactors distributed. One known means for trust management is the use of reputation systems. Reputation systems manage information about past behavior of interaction partners. Based on this information, individuals can get a clue how others might interact in the future. Reputation systems do not make expensive accountability measures (like, e.g., digital signatures under agreements made) obsolete, but aim to reduce the cases where expensive legal enforceability using these measures might become necessary. The state of the art and our research on trust management by *privacy-respecting interoperable reputation systems* is outlined in Section 3.2.4.

When introducing strong privacy mechanisms into Internet communities or collaborative networks, community selection and forming, i.e., finding potential collaboration and cooperation partners might become more difficult. Looking at real-world processes, mechanisms such as reputation management, advertising etc. might help. Thus, the decision to enter into a community mostly is made by getting to know it through advertising or rumors from friends (or from a friend of a friend and so forth) which would correspond to the reputation concept and which might be driven by social relationships forming social networks. Once more, such decision-making values depend on the particular contexts, i.e., factors like trust and dependencies play important roles. In Section 3.2.3 we outline the state of the art how *privacy-respecting establishment of collaborative groups* can be done.

3.2 Research results

3.2.1 Transparency support tools (Task 2.2.1)

Linkage control

Digital identities which represent people in the digital world are often linked with information about this very same person, e.g., social contacts or actions performed under that digital identity. In addition, this information can be further specified or extended by linking it with other data sources, e.g., other digital identities of the same person, and utilizing scoring models or other sophisticated algorithms which analyze the data.

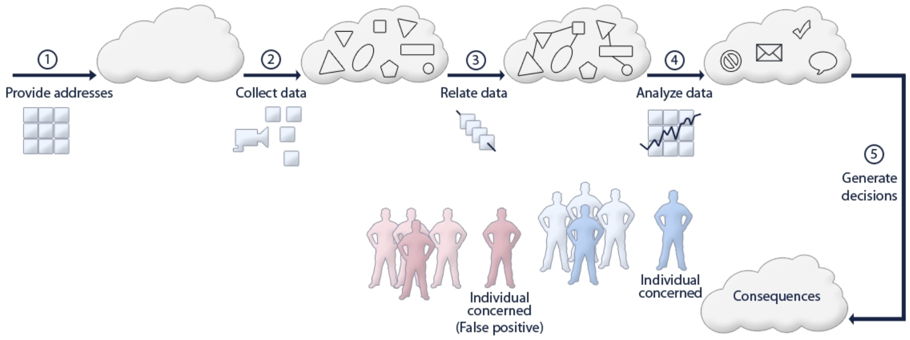


Figure 1: General model for enriching information

Figure 1 shows the typical data flow when enriching information for the purpose of generating decisions, as this is done many times a day in common data processing systems. For discussing linkage properties and objectives, it is important to make clearly visible who can access which data and perform which actions on these data. In all identified steps in the model work-flow presented in Figure 1, the actors contribute to some aspect of linkage and may have to be cautious to avoid undesired effects. In case of a mistake, it may be hard for individuals concerned to find the error and its cause in this work-flow and to achieve that appropriate corrective measures are being taken. As elicited in [Han08b], linkage control is the essence of privacy protection. Control can be based on transparency and checkability. This requires that the complex world of today's data processing with manifold actors has to provide all relevant information to check correctness and fairness of decisions.

The task of PE-IMS is to support an individual in managing their privacy and identity, and this requires in particular successful linkage control concerning the pieces of data relating to that individual. This bears several challenges: How to gather all this information which may not be public because of, e.g., trade secrets or national security reasons? Which granularity of information is necessary? How to establish information channels to the individual and his identity management system? How can the identity management system provide to its user a correct interpretation of these information? What are the options for the individual: How can he act and react upon the gathered information? What roles do or can third parties play, e.g., supervisory authorities, intermediaries, information providers, etc.?

In the first project year, a wide perspective was taken to derive the abstract model of linkage control which - from the author's point of view - is the objective for all kinds of transparency tools combined with identity management systems. This model is not limited on client-server communication, but is valid for all kinds of communication, e.g., peer-to-peer communication as tackled in the project. In addition the work has been aligned to WP1.3 where lifelong privacy aspects are elaborated. For linkage control (and transparency tools), the perspective of privacy throughout life is of major importance. Further work will concentrate on pragmatic solutions regarding transparency tools which obviously won't be able to implement immediately perfect linkage control, but at least will improve the current state significantly.

Transparency tools

One of our goals is to develop tools and concepts for increased transparency. As a first step to reach this goal and to get an idea of what has been done in the area and the current state of the art, a short survey of transparency tools for privacy purposes has been conducted. In this process we have also tried to find a way of categorizing these tools. A more in depth description of the survey and its results are presented in [Hed08]. Even though we realize the great importance legal, social and economical tools, frameworks and sanctions play in the transparency area[1], the focus of the survey has been on technical tools.

In order to limit the scope of the survey and to understand what we are examining we have defined what we mean by a transparency tool for privacy purposes. First of all, transparency as such can be required for more than privacy purposes, e.g., different types of audit and control to make sure that company finances are in order or that procedures and processes are managed and used in an appropriate manner and of course there exists tools to aid in those cases. In this survey we have limited ourselves to consider tools that have the objective to help the user to enhance her privacy. So, what is a transparency tool for privacy purposes then? The EU project FIDIS (`http://www.fidis.net`) has in its deliverable D7.12 [D 7] defined a concept called Transparency Enhancing Technologies (TETs). However, their vision on TETs is for tools that make it possible for individuals to assess how profiles will be used on them and to be able to judge how different actions will influence the outcome of this profiling. In our view this definition is too narrow considering the implications of the word transparency. Further, since we do not consider legal tools the definition is too wide in that sense.

In [Han07], Hansen presents a definition of transparency tools. This definition is, we believe, too narrow since it only takes into account the end user and not entities that act on behalf of a user or in the interest of the user to increase the user's privacy such as data protection officers.

Based on the definitions referenced above and on the classification on privacy mechanisms given in [SSA06] we would like to give the following definition on transparency tools for privacy purposes (please note that by a proxy acting on behalf of the user we also include organizations authorized by other entities than the user to protect the privacy interests of the user): A transparency tool for privacy purposes is a technological tool that has one or more of the following characteristics:

- gives information on intended collection, storage and/or data processing to the data subject, or a proxy acting on the behalf of the data subject, in order to enhance the data subject's privacy;
- provides the data subject, or a proxy acting on the behalf of the data subject, with access to stored data and/or to logic of data processing in order to enhance the

[1] Even though there exist technical tools for transparency we believe many of them require additional legal tools or technologies such as reputation systems and black lists in order to be fully effective. This is because there is limited use in getting the information if the person involved cannot act against the service if the promises are broken or her personal data are misused in some way.

data subject's privacy;

- provides counter profiling capabilities for a data subject, or a proxy acting on behalf of the data subject, in order to 'guess' how her data match relevant group profiles that may affect her risks and opportunities, implying that the observable and machine readable behavior of her environment provides enough information to anticipate the implications of her behavior.

For designing privacy enhanced systems it is helpful to have a classification that can be used in order to compare different tools or choose the right tool for the system. In [Hed08] we compared and analyzed tools based on the following characteristics.

- Possibilities of control and verification
  - "Promises": The user gets information on what the data controller promises to do or not to do with the data in the future.
  - Read only: The user or her proxy can get access to information on what processing the data actually has gone through up to a specific point in time and/or to the stored personal data itself in a read only manner.
  - Interactive: The tools have the ability to let the user or her proxy additionally to reading actively influence the stored data and/or the processing of the data in some way according to legal requirements or agreed on policies.
- Target audiences
  - Tools for Data Subjects are expected to have a high level of "user friendliness" and a high degree of automatization when it comes to interpreting policies or finding privacy violations. One would expect these tools to give advice on how to proceed or who to contact in case of privacy violations or questions.
  - Tools for Auditors/Proxies do not necessarily produce output that is presented or explained in a way that is supposed to be read or understood by non-professionals. These tools might give direct access to data or processes that are outside of what a Data Subject would be allowed to access or expected to handle or understand.
- Scope
  - Service Scope: Transparency to information stored and processed by a single service is given.
  - Organizational Scope: Transparency to information stored and processed by a single organization is given.
  - Conglomerate Scope: Transparency to information stored and processed by a conglomerate of organizations is given.
- Trust requirements
  - Trusted Server: The server environment is assumed to behave in a trusted manner.

  - Trusted Third Party: The solution requires that parts of the responsibilities and functions are taken over by an impartial third party component. .
  - Trusted Client: The client environment is assumed to behave in a trusted manner.
  - No trust needed: The solution itself is designed in such a manner that it, in some way, prevents (or makes it exceedingly hard for) the server and the client from cheating or misbehaving. This is achieved without the use of an external trusted third party to guarantee the trustworthiness of the solution.

- Information presented
  - Required information: The tool gathers and presents information that a service provider has to provide according to the Law (in a EU context this would, e.g., be national laws based on the EU Data Protection Directive 95/46/EC Art. 10 [Dir].
  - Extended information: The tool gathers and presents information given or harvested from the service provider that is not legally required but that increases the transparency for the user in a privacy context.
  - Third party information: The tool gathers and presents information given or harvested from other sources than the service provider that increases the transparency for the user in a privacy context.

3.2.2 Privacy measurement (Task 2.2.2)

Privacy measurement allows to assess the privacy that a communication system can provide or, more important in the context of identity management, to assess the privacy of an individual that acts in a communication system. Both approaches can focus on traffic and location data, which is for instance prominent for mixes, or focus on application data, for instance personal data that has been sent. A challenging question is currently how to combine the network layer with its communication data and the application layer. While privacy measures for communication systems provide usually an averaged assessment over all system states (or alternatively the privacy which is worst among the system states), privacy assessment for individuals is much more focused on a point in time. Still, the attacker dimension is usually aggregated as an average value. That is, due to a lack of knowledge about the actual attacker, the privacy that can be provided against each possible attackers (in a well-defined attacker model) is averaged and this average value is what the measurement results in. Again, the average value can be replaced by an extreme value, for instance the privacy that can be preserved against the most successful attacker. Though, replacing the average by an extreme value can be more or less appropriate depending on the use-case. The decision between the average and extreme value should be generalized to a continuous weighting that takes both values in consideration.

In order to perform a privacy measurement, a lot of data about the communication system and the events therein is required. The assessment only works as good as the quality and quantity of data is. The best assessment can be achieved with the data that is available to the real attacker, accompanied by the knowledge that this is in fact the attacker knowledge. With this data, the attacker can be simulated and the remaining privacy against this attacker can definitely be determined. But the comprehensive awareness about the attacker knowledge is in most cases unrealistic, it would be comparable to having a snapshot of the state of the world. Instead, privacy measurement has to assess the privacy on assumptions about the attacker knowledge. This seriously limits the validity of the measurement. An additional source of discrepancy between measured and real privacy are unrecognized errors in the data.

Privacy measurement has recently received a lot of attention. In the course of different traditions, several approaches have been developed for privacy measurement and how to use and extend existing privacy metrics. Compared to Task 2.3.1 our work focuses on validating and combining existing privacy metrics. We conducted a survey on the state of the art in this research field for FIDIS, a Network of Excellence (NoE) project of the EC (project no. 507512), and published in the FIDIS deliverable 13.6 on "Privacy modelling and identity". Fields our survey covers are formal methods, statistical databases, dataflow analysis in networks and information-theoretic approaches. Based on this survey and our publications during the last years years [SK03, Cla06, CS06, Cla07] we will work on further publications about privacy measurement. We started this already but the papers are currently under review.

3.2.3 Privacy-respecting establishment of collaborative groups (Task 2.2.3)

When discussing privacy issues related to group building processes, one needs to understand what collaborative groups are and how their establishment is characterized.

Collaborative groups exist with different characteristics. They differ in their objectives and kinds of interaction. Schlichter et al. [SKX98] describe three general types of groups: communities, social groups, and teams. It is not possible to give a clear, distinctive definition of a collaborative group. Generally speaking, collaborative groups can be described as: *A grouping of people developing, elaborating on, sharing, consuming, and organizing knowledge.*

Besides the differentiation between communities and teams as well as between the interaction modes collaboration and cooperation, collaborative groups can further be distinguished according to their size (number of members), their structure (e.g. what roles are used, substructures etc.), their envisaged life cycle (continuous existence, limited to the completion of a task etc.), and their envisaged communication approach (offline, online, hybrid).

When studying privacy issues of group-building processes as well as the requirements accruing from those issues, then considerations of the different phases of group development processes are necessary. Tuckman [Tuc65] distinguishes between five different phases: Forming, Storming, Norming, Performing and, Adjourning. Of course, there is no clear separation between the individual phases. Instead, seamless transitions as well as fallbacks occur between them. However, specification of requirements with respect to supporting privacy control highly depends on the actual stage of the group-development process. This is especially true for very dynamic groups where individuals may become group members and may also leave a group at any time.

These many factors make a general and unambiguous requirement analysis for privacy drivers in group-building processes nearly impossible. Further, in order to reasonably study privacy issues of group building processes, i.e. searching for and finding of potential group members, it is indispensable to study the initial situation for group development as well. Different possibilities of finding collaboration partners are outlined in [PBPH^{+}09]:

These heterogeneous situations make the research problem even more complex.

One of the main requirements in the field of user-controlled privacy and identity management is designing systems starting from maximum privacy. Thereby, most of the people would understand the latter as starting from full anonymity. Obviously, this cannot be realized when dealing with searching for and finding of potential members of collaborative groups. There should be information available, which allows for judging the competencies of a person as well as if her characteristics match the group's requirements.

Scenario of building a collaborative group

To realize the scenario of team building for work via a groupware, we assume that an organization (which might be an enterprise or a research association) provides a groupware to allow for managing profiles, support of communication, and coordination of work as well as for elaboration and central storing of documents. Further, we assume that one of the associates of that organization was commissioned a project for which she needs to set up a team. Efficiency of the new team depends on different factors, e.g., on the work conditions (common objective, specific goals, conflicts) and on the composition of the team (competencies and personalities of its members, roles, size and flexibility of the group), cf. [RJ08].

Search for potential team members can be performed using the information stored in the user profiles of the groupware. This information is composed data indicated by the user herself as well as gathered from previous activities the user was involved in.

In order to give the scenario a realistic and comparable touch, we give it particulars by placing it in a situation of organizing a conference. The groupware that is used by the conference organizers for coordination purposes related to the conference represents a web-based conference tool.[2]

Analysis of the scenario with respect to privacy issues

- During the *forming* phase, different steps are being accomplished. One of the first steps relates to creating a new conference entry within the conferencing tool by the initiator. Further, she will search for partners with according research interests who potentially join the PC. Typically, the PC is being constituted with people the initiator either knows personally, from their research work (papers etc.) or she gets to know by recommendation. A further possibility to find potential PC members could be using according directories, e.g., the database of the conferencing tool. The PC is being established by contacting the persons found and receiving their acceptance of joining the PC team. Thus, this step does not imply any privacy-related issues.

[2] Examples for such conferencing tools are: EasyChair (http://www.easychair.org/) and OpenConf (http://www.openconf.com/).

- *Storming* and *norming* can be observed with different steps of the PC members' work. One example is the creation of a text for the Call for Papers and the decision on according dates for submissions and the conference itself. Obviously, each PC member wishes to see her own interests reflected within that document. Since this may conflict privacy-related attitudes of the discussion participants, it seems reasonable to apply privacy-protecting mechanisms in such stages. However, full anonymous communication may hinder the creation process of the Call for Papers, i.e., within target-aimed discussions, it is necessary that given arguments are possible to be associated with the corresponding participant. Further, as within the performing phase (q.v.), being aware of a discussion participant's competencies fosters judgements of her contributions. According to this, providing full anonymity for the participants would hinder the processes. That's why, privacy-enhancing identity management is a recommended means to be applied.

  Further, the members need to agree on a certain content that meets all interests. So, this step may elicit the requirement of supporting anonymous voting. Anonymous voting may also be applied to the situation of deadline extension for paper submissions. Taking a clear stand in this context might allow to draw conclusions with respect to personal attitudes. Thereby, it must be ensured that only PC members submit their votes as well as they can do this only once, but anonymously.

- The announcement and distribution of the Call for Papers to people working in the according research field is counted to the *performing* phase. Here, it is not necessary that PC members get to know the identities of potentially interested contributors of papers. In practice, the Call for Papers is published via established channels whereby recipients remain anonymous.

  The step of assignments of papers to reviewers requires different privacy-related considerations: First, the authors' interests have to be protected. This might mean, reviewers have to do their work unpersuaded and must not be enabled to be biased because of knowing the author's identity. Second, the reviewers' identities should be hidden to allow for equal rights for all involved parties. However, at the same time it must be ensured that the reviewers possess the according competencies to judge the authors' papers. Also, it must be ensured that conflicts of interests (such as author and assigned reviewer are one and the same, they know each other very well, or they have a conflict-ridden relationship to each other) are circumvented. Further, after submission of the review, an author should have the possibility to discuss issues of the review with the according reviewer. This needs to be supported without disclosing the real identities of the parties.

  Discussion on papers' acceptance can potentially be performed without revealing the real identities of the involved PC members. However, final decisions must regard the PC members' proven competencies related to papers, which is subject to acceptance decisions.

- Within the *adjourning* phase that follows the actual conference, different final activities have to be accomplished. One of this might be updating the database of the conferencing tool with information about people who participated in the PC or giving ratings to the PC members reflecting the quality of their work.

Recommendations. The analysis described above has shown that several issues exist connected to privacy interests of parties involved in the conference scenario. In the scenario, we assumed a rather traditional approach where PC members typically know each other. In order to allow for a scenario that stresses the privacy interests more than the usual procedure, it would be necessary to include the possibility to use pseudonyms in the design of the conferencing tool. Thereby, people should be enabled to use different pseudonyms for different contexts. In this case, the scenario within the application (i.e. the conferencing tool) should be partitioned into several privacy-related contexts[3], cf. [BDF$^+$05]. So, any time a user changes the context, the system would allow to switch to another pseudonym. Examples of such privacy-related contexts would be communication with other PC members, discussions with authors about contents of papers (where authors should not be able to link the reviewer to her real identity), pseudonymous petitions, and certain decision making processes (e.g. adjudications or exchange of personal opinions).

If the scenario is partitioned into different contexts then it must be ensured that the users do not arbitrarily change their pseudonyms, i.e., for some contexts, every time a user enters one of these contexts, she has to appear with one and the same pseudonym again to be recognizable, while this is not necessary for other contexts. A possibility to allow for the implementation of those constraints represents the use of policies, which can be used to regulate such required behavior.

The use of policies is also required for access control of resources, e.g., submitted papers, submitted reviews, personal information of PC members etc. Because of the recommendation to provide the possibility of using multiple pseudonyms, traditional access control mechanisms like access control lists (ACL) or simple role-based access control (RBAC) cannot be applied. Therefore, we recommend to introduce mechanisms based on anonymous credentials and data access policies as proposed in [FWBBP06]. Thereby, users are issued anonymous credentials, which, in conjunction with policies, are used to prove the right of the credential's owner to access certain resources. Additionally, anonymous credentials can be used to certify one's competencies in a particular research field, which would, e.g., be required within the decision processes regarding acceptance of papers in that research field. Another use case of anonymous credentials is anonymous voting. Thereby, support can also be given to ensure that the voter is entitled to participate in the voting. Further, special credentials allow to fulfill the requirement that a vote is given only once by the same party. Another use case of policies and credentials in the given scenario represents the submission of reviews. With help of policies and credentials, the process of verifying that only entitled reviewers respectively her proxies can submit reviews can be supported.

Efficient work in teams is intrinsically tied to trust between the team members. However, in privacy-respecting environments where users are enabled to learn only a limited feature set of other persons, mechanisms for establishing trust among the members are required. Thus, we suggest to consider the implementation of according reputation mechanisms.

[3] We call this procedure *Intra-application Partitioning (IAP).*

3.2.4 Trust management by interoperable reputation systems (Task 2.2.4)

Whenever people interact with each other they want to know what to expect from others and then want to trust in the fulfillment of their expectations. Social scientists and theoretical economists model the problem whether two interaction partners should place trust in each other as a so-called trust game [CW88, Das00].

Internet users often only interact once with each other. To help new interaction partners to estimate the others' behavior reputation systems have been designed and established to collect the experiences former interaction partners made [RKZF00]. Interactions can have various forms, they might reach from just reading content others wrote to playing interactive games with each other.

As an assisting system for interaction systems a reputation system tries as well to increase the interaction system's security as it has to be multilateral secure itself. Our research in this field will be published in [Ste08].

Multilateral security and reputation systems

Multilateral security. When interacting with others, users necessarily have several (sometimes contradicting) security requirements. Here multilateral security means providing security for all parties involved, requiring each party to only minimally trust in the honesty of others [RPM99].

Interactions usually consist of several actions depending on each other. These actions are usually transmitted as distinct messages. For a single message security requirements of its sender and recipient(s) are well studied, e.g. in [WP00].

Interaction partners necessarily have security requirements in common concerning the content of the message. These requirements can be fulfilled by technical measures only to the extent the interaction partners cooperate and behave as expected. This means, interaction partners typically have an expectation regarding the others' behavior that these might fulfill or not.

For this reason the security requirements regarding the content of the message have to be reformulated in comparison to [WP00] as outlined in [BPHL$^+$07a, Ste08].

An interaction usually consists of numerous individual but correlated actions. Correlating actions lead to new security requirements as outlined in [Ste08].

Reputation systems. A reputation network is a social network that links entities (possibly pseudonymously) to each other and allows them to interact and exchange information with each other. Entities can learn possible interaction partners' reputation from former interaction partners or other entities who observed the possible interaction partner. In social sciences this is called the *learning mechanism* of the reputation network [BR01]. Entities may also control others in the reputation network by spreading information about the entities' former interactions. In social sciences this is called the *control mechanism* of the reputation network [BR01].

Both entities and interactions within the reputation network can be reputation objects. Entities and non-completed interactions are *dynamic reputation objects* while completed interactions are *static reputation objects.*

Reputation systems assist reputation networks technically. For interactions within the reputation network we assume different interaction systems to be in place (e.g., simple e-mail, file sharing, community systems). Then there are the following five components of a reputation system.

- *Rating algorithm* of a rater. This implements the control mechanism of the social network.
- *Reputation algorithm* for reputation update. The reputation system updates the reputation of the reputation object from the ratings received.
- *Propagation of reputation and ratings* for evaluation of a reputation object's reputation (see below),
- *Storage of ratings and reputation.* After the creation of reputation it has to be stored somewhere. Reputation might be stored *centralized* at reputation servers designated for this purpose, *locally* at the device of the user whose pseudonym received the reputation or *distributed* at the devices of other users.
- *Evaluation of a reputation object's reputation.* All members that influence the reputation of an object by their ratings, additional trusted third parties, the reputation object itself and possible future interaction partners might evaluate a reputation's object. Here the reputations received need not be the same for every evaluator. The reason for this is that the selection of ratings used for the evaluation depends on both the information flow of ratings in the reputation network and the trust structure on the reputation network, i.e. how evaluators trust in ratings from other members. Those who rate need to be trusted in giving a correct rating which is in line with their view on a specific interaction. The reputation selection for evaluation can be *global* or *individual.*

To find design options for these components one has to consider several security requirements. Beneath the security requirements for communication based on the functional requirements of the learning and control mechanism of the reputation network new security requirements for reputation systems can be identified.

- *Bona fides of ratings and reputation.* If it would be possible for all members of a reputation network to observe an interaction and if all of them would give the interaction the same rating this rating would have *objective bona fides.* But most ratings depend on subjective estimation of the interaction partners or observers at a certain point in time. As a special action a rating has *subjective bona fides* for an observer if it corresponds to her expectation of the interaction. Accordingly a reputation has subjective bona fides if it is created by bona fides from ratings done by bona fides.
- *Fairness of the underlying game-theoretic trust game.* A reputation system is fair if every authorized user has the same possibilities for rating an interaction partner. The authorisability in the reputation system has to follow the control mechanism of the reputation network. Only users that gave a leap of faith to interaction partners should be able to rate them.

- *Completeness of reputation.* Members of a reputation network expect to receive as much information as possible from interactions performed in the reputation network. This needs the willingness of authorized interaction partners to rate each other and the willingness of all members to distribute reputation in the reputation network.

- *Persistence of reputation objects.* To help the control mechanism to be employed longevity resp. persistence [RKZF00] of members as reputation objects has to be realized resp. the binding of reputation to them. This can be done pseudonymously.

- *Absolute linkability of a user's membership in a reputation network.* To prevent a user from leaving a reputation network with a bad reputation and re-entering it with a neutral reputation membership actions of the same user in the same context have to be absolutely linkable.

The rating and update of reputation has to follow specific rules fixed by the system designer. These rules usually depend on the application scenario and have to fulfill sociological and economic requirements. We abstract here from concrete functions to allow a universal design interoperable with various IMS and application scenarios.

By actions in the reputation network no other requirements on interactions should be affected. This needs unlinkability of actions and pseudonyms of the same user in different interaction systems and the reputation network as well as providing anonymity to her. By interoperability with a PE-IMS a user can at least use different unlinkable pseudonyms for every system she users for interaction.

In the first year we developed a multilateral secure system design for two scenarios as we will outline in the following.

Rating interactors in community systems

The first scenario/infrastructure we chose to test a multilateral secure reputation system for rating interactors is a centralized community system where interactions between members take place via a central server where they are stored and globally available. We chose the web forum software phpBB[4] for our implementation. Our system needs an identity provider to assure accountability of the pseudonyms used. The credentials and the required functions for handling them were implemented using the Idemix-Framework[5], which is written in Java. Our prototype does not use a PE-IMS yet but uses the authentication methods of phpBB. Therefore the registration process takes place simultaneously in the phpBB community and the reputation system. The phpBB community could be used as usual, but the member can call the reputation functions within the phpBB interface that have been extended for this reason. For details we refer to the related publication [PS08a].

Rating web content in the semantic web

The second scenario/infrastructure we chose is the rating of arbitrary web content in the Web 2.0. This system design will also be published in [SGM09]. Beneath the reputation

[4] `http://www.phpbb.com/` (last visited Jan. 09)

[5] `http://www.zurich.ibm.com/security/idemix/` (last visited Jan. 09)

	User-independent	**User-specific**
Application-independent	Talks, campaigns, tutorials	Individual advice from a Privacy Commissioner
Application-specific	Privacy disclaimers on web sites	Feedback from web site's policy evaluation (e.g. Privacy Bird)

Table 2: Examples for dimensions of privacy-awareness information

system itself the interoperability with a PE-IMS and a PKI is needed as infrastructure. We validated our system architecture by a first prototype based on the Java Runtime Environment 1.5. The Java Security API is used to realize the necessary cryptographic primitives, i.e. cryptographic hash functions and digital signatures. Our design is open for interoperability with a PE-IMS as outlined above but currently does not implement it. The core component is implemented by a Java application to be locally run at the user's device. This makes it usable for the average Internet user as we evaluated in a small user test.

3.2.5 Privacy awareness (Task 2.2.5)

The objective of our work is to research privacy awareness especially with regard to collaborative applications from a technical as well as from a human perspective. Therefore we want to design tools to support privacy-awareness and also study the influence of privacy-awareness support in users' behavior.

Awareness is based on an individual's perception, cognition and attention on certain physical as well as non-physical objects. The state of being aware of something fades away as soon as there is no longer any stimulus present. Information from the environment or from other individuals constitutes such stimuli.

There are two main approaches for content and representation of information that serve as a stimulus for a user's awareness of privacy: the user and the application. With the first approach privacy-awareness information is of general nature, i.e., independent from a users' individual preferences and independent from a particular application. With the second approach privacy-awareness information is geared personally to a user or to a specific application. Examples for these different dimensions of privacy-awareness information can be found in Table 2.

The privacy paradox

Privacy awareness enables users to make informed decisions and should lead to less user actions that endanger the user's privacy unintentionally. Consequently, it can be assumed that users who state the intention to protect their personal data, i.e., who can be considered privacy-aware, will act according to their statements if they have the choice between different options for action. However, several behavioral studies show a contradictory finding (e.g. [SGB01, NHH07, JPJ05, CS05]).

An online shopping experiment compared users' self-reported privacy preferences with their actual self-disclosing behavior and found out that a majority of the test participants – regardless of their formerly stated privacy attitudes – disclosed a large amount of personal information [SGB01]. Similar results are shown in other study about users' intentions and behaviors towards privacy, where the participants provided significantly more personal data than they claimed that they would beforehand [NHH07]. Within this study the researchers also tested whether the perception of risks is more salient and has a negative influence on customers' stated intentions when they are asked such a question in general, whereas this is not the case in specific real situations when customers actually decide to disclose data. This hypothesis was supported by the results of the study.

The contradiction between attitudes towards privacy and actual behavior, identified in the cited studies is labeled the privacy paradox [NHH07]. Although these findings derive from studies that were designed with regard to e-commerce applications, first results from [AG06] indicate the existence of the privacy paradox in online social networking applications. Acquisti and Gross found out that one group of privacy-concerned users simply does not join in online social networks, which is not surprising. However, those who have a high level of privacy concerns, but are members of an online social network, share nearly the same amount of personal data (e.g. birth date, sexual orientation or personal address) as other members of the network. As discussed more detailed in [PÖ9], a reason for the privacy paradox is a *balancing of different values.*

General requirements on tools to support privacy awareness

Privacy awareness is important for users in order to make informed decisions about the disclosure of data. Especially for users who are not familiar with privacy and privacy settings, tools to support privacy awareness can be an *assistance* to learn about how to set privacy preferences and what possible consequences the disclosure or non-disclosure of data may have. Though, the presentation of consequences can never be a perfect prediction of reality, since it is not possible to consider all external incidents in advance. Tools to support privacy awareness also encompass means that provide *feedback* to users about their current level of privacy within the application. However, such a "privacy-o-meter" can be very misleading and give a false feeling of privacy depending on the assumptions of the model and the accuracy of calculations in the background, which again will not consider influences from outside the application. As *reminder* tools to support privacy awareness helping people not to forget about their privacy preferences by indicating to the users which personal data they are going to disclose to whom. Therefore the user either can explicitly state her/his preferences in advance and the tool matches user's preferences and policies of a communication partner automatically (e.g. PrivacyBird) or else the tool just informs the user what data will be transmitted. In the latter case the user needs to check herself/himself in thoughts whether she/he accepts the disclosure or not. This feature is for people who already have some knowledge about privacy and can think about possible consequences themselves.

For the design of tools that support users' privacy awareness, a number of general requirements emerge and are explained in the following. Ambivalences, which ensue from the demand for a high flexibility of tools, user-control and freedom of choice for the user on the one hand and strict definition of rules for implementation on the other hand, are discussed.

- *Measure user's privacy attitude.* In order to "remind" users about their privacy attitudes in specific situations, their general attitudes have to be known by the support tool. There are two ways to capture users' privacy preferences: (a) ask them directly or (b) gather preferences from observation of actual behavior. The latter option has at least two problems. First, monitoring of users' behavior might be privacy-invasive itself and, second, the privacy paradox describes the gap between attitude towards privacy and behavior. Hence, drawing conclusions from monitored behavior would simply not help. Asking users directly means in fact to let them customize their tool for privacy-awareness support. The challenge here is to motivate users to configure and to change preferences, particularly since usually users rarely customize their preferences but rather use default settings [Mac91, GAHJH05]. Cognitive science refers to this phenomenon as the "status quo bias".

- *Understandable for users.* The choice of words and descriptions needs to be understandable for ordinary users, not only for computer specialists. It is not sufficient to rely on experts' opinion about what might be useful to display and how to inform users. As pointed out in [AS01], it is important to identify and consider users' perception, understanding and needs for designing usable applications. The majority of users are not experts and the level of technical knowledge differs among them.

- *Consider cognitive boundaries.* The concept of "bounded rationality", which is well known in cognitive science, signifies the limited ability of individuals to acquire, process, and memorize information [Sim82]. That is, even if users would theoretically have all privacy-relevant information available, they will not be able to use all this input for making a rational decision, however they apply a simplified mental model. When designing tools to support privacy awareness this needs to be considered and opportunities have to be researched how to present data to users in a way that they are able to handle it cognitively.

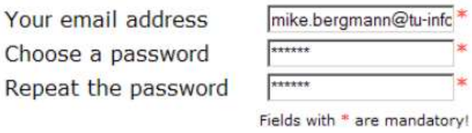
- *Tailored to concrete situations.* Tools to support privacy awareness should influence users' behavior in concrete situations and therefore need to be user-specific and application-specific. Presentation of information should depend on the current context, i.e., the user's task, kind of information, recipients, usage, etc. This means either a rule set of all possible contexts has to be defined beforehand by the system's designers or users need to configure their personal sets of contexts, which means making an additional effort for them.

- *Offer support, no assumption of responsibility.* Tools need to be designed in such a way that they offer support to users. They should not convey the impression that they fully protect users' privacy according to their preferences or that there is no longer any need for users to be aware of privacy and to take care for themselves.

- *Performance.* It is essential that tools or features for privacy-awareness support do not decrease performance of the primary application to a perceptible extent, since users will not accept long delays. This is documented for usage of web sites [GHMP04], anonymization services [Koe06], and it is assumed to be true in terms of privacy-awareness support as a secondary feature as well.

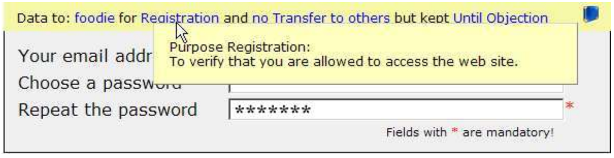
A first test about the awareness of privacy policy statements

Within the first year we made a privacy-awareness test that was also published in [Ber09]. We rephrased the previously given definition of privacy awareness in a more concrete way in order to make it operationalisable. Thus, in the test we studied *the ability of the user to reflect the privacy policy statements of the communication partner regarding purpose binding, transfer assertion and retention period applied for a data disclosure.*



Your email address mike.bergmann@tu-infc *
Choose a password ****** *
Repeat the password ****** *
Fields with * are mandatory!

Figure 2: Conventional interface for online forms



Data to: foodie for Registration and no Transfer to others but kept Until Objection
Purpose Registration:
To verify that you are allowed to access the web site.
Your email addr
Choose a passw
Repeat the password ******* *
Fields with * are mandatory!

Figure 3: Enhanced interface – information about the privacy policy nearby the data to disclose

The participants of the test were randomly assigned either to the *Control Group* G_{NoPet} using the conventional web forms for data disclosure (see Figure 2) or to the *Experimental Group* G_{Pet} using our enhanced interface presenting the details regarding the privacy policy of the service provider (see Figure 3). One of the questions the test should answer is: *Does the user really perceive the privacy policy statements, presented in a superficial manner such that we could achieve an increased privacy awareness?*

The survey was announced to a broad audience in various online forums. Counting the participants using the start button to start the survey, we got 618 participants. Referring to a classification by Alan Westin [KC05] in our sample about 18% of the participants belong to the share of people who are *Unconcerned* about their privacy, about 48% were *Pragmatists* and about 34% we count as *Fundamentalists*. In this context, a sub-question will be how the perception of the privacy policy differs among the various classes. Our hypothesis is that privacy fundamentalists and pragmatists appreciate the enhanced presentation form, while the unconcerned users still ignore it.

Our test showed this is indeed the case. Figure 4 shows the mean value for the *privacy-awareness index* per group and class. Figure 4 also illustrates that the ability of the participants of the experimental group G_{Pet} to reflect the main preferences regarding the privacy policy, stated by the service provider, is higher. The significance of this outcome we prove with Pearson's chi-square test. The results are stochastically independent with the probability of approximately 0.995%.

We have shown that the proposed approach for presenting information related to the privacy policy of a certain transaction does significantly help the user to perceive the essential privacy preferences, like purpose of data usage. The effect was observed for all classes, but with most success for pragmatists. Our hypothesis was confirmed. The effect is even excelling our expectations because the increase of the privacy-awareness

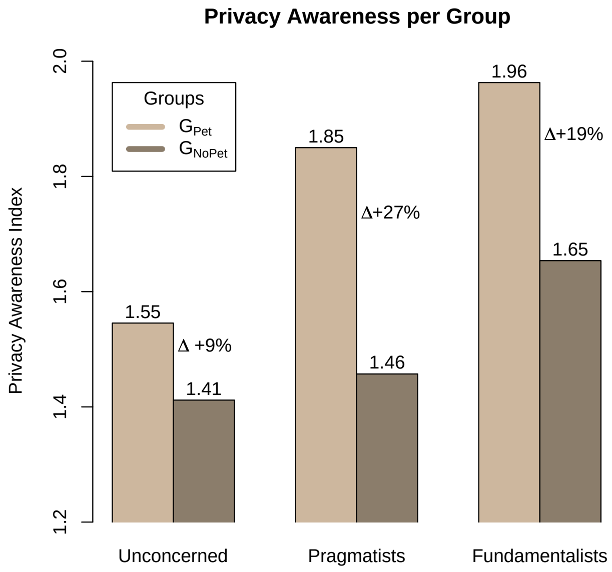


Figure 4: Contrasting experimental and control group regarding mean *privacy-awareness index*

index was also observed for the class of fundamentalists. Initially we expected that the fundamentalists read and perceive the preferences of the privacy policy anyway. The increase of the privacy-awareness index may be due to the mismatch between stated vs. observed behaviour (cf. 3.2.5). So we may conclude that the usage of enhanced interface pays off for all classes of web users.

3.3 Future research

Our goal for the remaining time of PrimeLife is to implement and test our research results in practical tools. From both the part of the WP focusing on privacy (privacy measurement, awareness, transparency) and the part focusing on trust (group building and reputation) at least one tool will be implemented and tested within Activity 1. For the latter part this tool will be from the field of reputation because there exist already two stand-alone tools developed within the first year. For the first part it will hopefully be possible to develop a tool that integrates aspects of privacy measurement, awareness and transparency.

Beneath the practical part on tools and the separate research aspects in every research task there is also the need to integrate aspects of the different parts. The main focus here lies on the the contradiction of privacy and trust and possible solutions.

3.3.1 Transparency tools (Task 2.2.1)

In the next two years we will, based on the knowledge gained through this work, try to implement a general transparency tool that will give the user more control over and knowledge about his/her data than is given by the tools implemented today. This work aims to try to evolve the implemented transparency tools within the PRIME prototype towards being interactive and to improve the other aspects of the transparency part of the prototype.Some of the main challenges that will be addressed in this process are:

- How to present and handle data relating to more than one person.
- How to achieve secure and privacy friendly access to personal data for data subjects. This also includes issues regarding proof of ownership of pseudonyms.
- How to harvest and present information for one data subject using different pseudonyms in the same context in a privacy friendly manner particularly regarding linkability issues.
- How to log and present privacy related events on the services side in a privacy friendly and secure manner. item How can the above mentioned challenges be solved in such a manner that they can be used by data subjects as well as by proxies in a secure and privacy friendly manner.

A first step will be to make the prime prototype interactive regarding data access in a service scope and to implement a first version of secure privacy friendly eventlogging. The main component in this effort will be the PRIME DataTrack together with an extension on the server side functionality in order to support the DataTrack regarding log functionality.

In addition we will work on integrating further information channels, e.g., for a notification of privacy and security breaches or vulnerabilities which could be performed by vendors, official bodies, NGOs as well as groups of active peers. All in all we will elaborate further mechanisms which facilitate linkage control by the user: This comprises both mechanisms supporting transparency of what is happening with the user's data or what may affect him/her later on and mechanisms supporting direct influence and control by the user.

3.3.2 Privacy measurement (Task 2.2.2)

In the last years, measurement aspects of anonymity and privacy have been investigated mainly with respect to single layers. That is, much research has been done with regard to metrics for mix-based anonymising techniques for the network layer. Similarly, various approaches for metrics have been developed for the application layer especially with respect to privacy-enhancing identity management systems as well as with regard to statistical databases. However, only few research has been done with regard to metrics for a combination of both layers. This research deals with the combination of both aspects and these results base on experiments with limited scopes.

In our our future research, we will further the understanding of aspects regarding the combination of layers. We will investigate relations and dependencies between network layer and application layer by developing models, which explicitly include both layers. Thereby, we aim at designing the models in such a way so that we can derive inter-layer dependencies directly from the model. Findings from this work may then lead to a better understanding of more comprehensive privacy metrics.

Another research direction will be to investigate effects of different attribute types on privacy metrics, especially with regard to non-static attributes. So, attributes may change more or less often over time, and prediction of changes may be more or less accurate, but could as well be rather impossible. Privacy metrics have to be designed in such a way so that these dynamic developments can be adapted.

3.3.3 Privacy-respecting establishment of collaborative groups (Task 2.2.3)

The work on this task documented within this deliverable gives a first overview of the requirements and recommendations with respect to privacy-respecting design of applications supporting collaborative groups.

Typically, the scenarios are not as simple as the conference scenario presented here. For instance, one may consider a situation where people get the possibility to decide about joining the PC team based on getting to know about the particular competencies and characteristics of the people being contacted to join the team. This would imply that those attributes need to be published in a way that the identities of the corresponding people remain hidden. So, while the analysis presented for the conference scenario was performed as a general approach, a more fine-grained analysis should be carried out, e.g. by applying the model of multilateral interaction (MLI) environments, which is being developed within Work Package 1.2 of the PrimeLife project. First ideas and thoughts related to the research field of MLI environments had been published in [BPHL$^+$07b, BPHL$^+$06]. Also, the analysis has to explicitly consider differentiation between privacy concerns within a group, i.e., towards other group members, and outwards, i.e., protection of group matters from external influences/privacy threats. So, future work needs to contemplate these different aspects as well.

Further, the analysis as well as the results concentrate on just one point in the landscape of collaborative groups – a particular type of team, namely the program committee of a conference. However, in order to get an overall picture of requirements related to privacy issues of collaborative groups, further scenarios focusing on other characteristics and features have to be analyzed as well.

3.3.4 Reputation (Task 2.2.4)

The concept how to design reputation systems in a multilateral secure way becomes more and more important with the growing number of applications.

Within the first year we developed and implemented technically two different types of interoperable reputation systems. In future work we will extend our systems to find a compromise between the different design options presented in the systems. Especially both rating web content and web users should be possible for numerous applications like the scenario of trusted content in WP1.1.

Our future research will also concentrate on different forms of interactions systems and the interoperability arising between them. Our focus lies on authentication methods using a PE-IMS like PRIME. Furthermore, we intend to integrate our systems with evolving user-controlled PE-IMS instead of a separate program. Here we will also cooperate with WP1.1.

3.3.5 Awareness (Task 2.2.5)

As next steps of our work we concentrate on the human factor and prepare a systematic study about the influence of privacy-awareness information on user behaviour.

The psychological theory of the *cues-filtered-out approach* is a topic of research in relation with computer-mediated communication environments [SK86]. In this scope, we hypothesize that due to missing social and context cues, computer-mediated communication, e.g. via web forum, facilitates the disclosure of personal data without taking into account the potential audience. Considering also the expectations for disinhibited behavior due to the lack of context cues, users may be very frank about telling personal details in computer-mediated communications and at the same time they tend to express insulting statements.

A tool to support privacy-awareness aims at aligning users' perceived privacy with their actual privacy. The research question we want to answer is whether and how such a tool influences users' behavior in an online environment that enables communication among members of a community, especially with regard to the disclosure of personal data. The results of this study constitute a valuable input for the further specification of requirements and the design of tools to support privacy awareness.

Chapter 4

Privacy of data (WP2.3)

4.1 Introduction

Data holders are increasingly finding it difficult to produce anonymous information in today's globally interconnected society. The technology advancements leave the information vulnerable to inference and linking attacks, meaning that from data that seem anonymous and from available public data, an adversary can make some inferences about sensitive data. This is possible because released information often contains other data that in combination can be linked to publicly available information to re-identify users. Existing approaches that are based on data distortion or disturbance do not satisfy the requirements of novel scenarios, where data must be made available subject to the constraint that data themselves must be truthful while preserving the privacy. In fact, data quality is of critical concern to organizations since data and information become key strategic resources and therefore poor quality can have significant consequences on the ability of organizations to act effectively.

The focus of Work Package 2.3 is then represented by the consideration of large data collections that contain sensitive information on citizens. The overall goal is the definition of novel metrics and techniques able to support the management of privacy requirements, at the same time offering a significant degree of utility in access to the data. The investigation of these topics in PrimeLife has two roles: on one hand, it can produce concrete techniques for the protection of personal information, identifying the amount of exposure deriving from access to the data and proposing approches able to satisfy the stringent privacy requirements that the project wants to support; on the other hand, the identification of metrics and techniques can provide input on the definition of components in the policy language able to express user preferences on the processing of their data (e.g. users can express the preference that their data when collected do not have to expose the user profile in groups with cardinality less than k). The availability of relational database technology is assumed for many of the scenarios considered in the Work Package. Work Package 2.3 is organized in the following three tasks, focusing on specific aspects of the privacy problem.

- *Task 2.3.1 Privacy assessment and privacy metrics* will define metrics and measures for formally describing the level of privacy protection that is guaranteed on a data collection.
- *Task 2.3.2 Techniques for enforcing data privacy* will focus on the definition of techniques for protecting data privacy according to some constraints that must be satisfied by the data themselves.
- *Task 2.3.3 Efficient organization and access to privacy-preserving data collections* will investigate the efficiency aspects related to the techniques enforcing data privacy, which are fundamental aspects to take into consideration to guarantee the applicability of the data protection techniques in very large data collections.

In the first year of the project, Work Package 2.3 has mainly developed novel solutions for enforcing privacy constraints in mobile networks and privacy requirements/constraints within business applications. In the following, we briefly describe these research results and highlight a number of possible directions as future work.

4.2 Research results

4.2.1 Privacy assessment and privacy metrics (Task 2.3.1)

The work in this task focused on the analysis of current privacy metrics for data collections and the design of novel solutions supporting privacy requirements in different scenarios, including mobile networks. The long-term goal is to identify techniques able to support data release for analysis without violating user privacy. Within this task, three main lines of investigation have been pursued: *i)* the analysis of privacy metrics from three angles (i.e., empirical research regarding privacy attitudes, privacy impact assessments (PIA's), and research regarding profiling), analyzing the state of the art in order to provide direct input to the lines of investigation in this task and in the other WP2.3 tasks; *ii)* the development of a solution for enforcing privacy constraints in mobile networks; and, *iii)* the consideration of privacy requirements within business applications.

Privacy assessments

Privacy Risk Perception. The data used for constructing an instrument for measuring Privacy Risk Perception (PRP) was collected as part of the PRIME project. An online survey was organized, with a questionnaire addressing demographics of the respondents. The respondents were also randomly assigned to answer questions about one out of four specific online contexts: e-government, online shopping, online banking, and online travel booking.

Privacy Risk Perception of the respondents was measured by seven items on a five point scale as part of the question: 'How concerned are you in general about the following possible consequences of abuse/misuse of your personal information?' followed by a list of concerns, like: 'threat to your personal safety', 'loss of personal freedom', etc. The respondents showed to be most concerned about invasion in their private sphere. They were least concerned about the possible threat to their dignity and about receiving an unjust treatment as a result of abuse or misuse of their personal data.

Moderate to high, positive, and significant correlations were found between the items. There are positive relationships between the specific privacy risk perceptions. This indicated that the items could represent a common concept. A principal component analysis was conducted to test this assumption. One component was found, indicating that these seven items together measure the same latent construct: privacy risk perception. The mean of privacy risk perception of the sample showed that on average, the respondents are neutrally concerned about privacy threats.

Privacy Impact Assessments. Based on the European Data Protection Directive (95/46/EC) and the corresponding national data protection legislation, general organizational measures are in place. A number of these measures ensue from the applicable European or national legislation. Balancing the legal obligations of the government, and the privacy protection requirements of its citizens, is an ongoing challenge. One of the tools to support this effort is the Privacy Impact Assessment. A Privacy Impact Assessment (PIA) is an analysis of how information is handled: (i) to ensure handling conforms to applicable legal, regulatory, and policy requirements regarding privacy, (ii) to determine the risks and effects of collecting, maintaining and disseminating information in identifiable form in an electronic information system, and (iii) to examine and evaluate protections and alternative processes for handling information to mitigate potential privacy risks.

Profiling. Behavioural Biometric Profiling (BBPs) and Transparency Enhancing Technologies (TETs) were studied in order to see how data controllers pick up on a variety of behavioural patterns, 'leaked' by citizens in their everyday lives, to model and predict their behaviour. In an Ambient Intelligent environment such pattern recognition would allow for massive group profiling, inferring a great many profiles that entail knowledge about health risks, earning capacity, life-style preferences, criminal intentions. The research question in this project was how transport data can be modeled and profiles can be generated to predict behaviour and preferences. Compared to the work in Task 2.2.1, the focus here is on providing direct requirements to the definition of novel privacy metrics.

Data Mining Without Discrimination. In this project the extent to which legal and ethical rules can be integrated in data mining algorithms to prevent abuse was studied. For this purpose, data sets in the domain of public security are used, made available by police and justice departments. The focus is on preventing that selection rules turn out to discriminate particular groups of people. Key questions are how existing legal and ethical rules and principles can be translated into formats understandable for computers and in which way these rules can be used to guide the data mining process. Furthermore, the technological possibilities are used as feedback to formulate concrete directives and recommendations for formalising legislation. Contrary to previous attempts to protect privacy in data mining, this research focus is not on (a priori) access limiting measures regarding input data, but rather on (a posteriori) responsibility and transparency.

Instead of limiting access to data, which is increasingly hard to enforce in a world of automated and interlinked databases and information networks, rather the question how data can and may be used is stressed.

Privacy in mobile environments

The second contribution is related to the exploitation of the k-anonymity privacy metric and multi-path communications for protecting the users's privacy while maintaining data integrity and accountability. Current privacy protection systems are focused on preserving users from untrusted service providers and are based on the assumption that the mobile network operators are trusted. Some current solutions use k-anonymity to protect the users during the communications with the service provider and consider the mobile network operator as a fully trusted party. Since the mobile network operator has however access to precise location and traffic information for each user, the goal is to develop a solution that can be applied also when the mobile network operator is honest but curious. This problem along with a possible solution have been presented in a publication [ASJ$^+$08].

The proposed approach builds on the concept of k-anonymity in the context of network communication, where a user is said to be network k-anonymous, with k the privacy preference of the user, if the probability of associating the user as the message originator is less than or equal to $1/k$. Such a network k-anonymity concept has been extended to hybrid mobile networks that are based on mobile parties communicating through wireless and cellular protocols to access services, either co-located in the cellular network or in the Internet. The entities involved are: *1)* mobile users that carry mobile devices supporting both GSM/3G and WiFi protocols for communication and that may request services to providers available over the network; *2)* cellular network (and corresponding mobile network operators) composed of multiple radio cells, which provide network access and services to mobile users, acting as a gateway between mobile users and service providers; *3)* service provider that provides on-line services to the mobile users and collects their personal information before granting an access to its services. In such hybrid networks, users can simultaneously create WiFi point-to-point connections, join the cellular network, and access the Internet through their mobile phones.

The proposed approach has been developed by also having in mind the necessity for mechanisms to make the users accountable for their operations. In fact, some current solutions can be abused or lack economic incentives due to the lack of user accountability. Service providers are often reluctant to adopt privacy solutions that completely hide the users and do not enable any form of accountability. The basic idea of the proposed solution is that using a multi-path communication paradigm, a mobile user can achieve network k-anonymity by distributing, using WiFi network, different packets of the same message to k neighboring mobile peers, which then forward the received packet through the cellular network. This scheme achieves k-anonymity because the mobile network operator is not able to associate the users' data flow with fewer than k peers. A separate accounting mechanism can verify that the packets are legitimate. For instance, one approach is to have the data flow encrypted with a symmetric key shared between the requester and the service provider. This would assure accountability, data integrity, and confidentiality. In addition, it will prevent the abuse of anonymity while providing the economic incentives to deploy anonymizing schemes.

While most of the existing privacy solutions and anonymous routing algorithms heavily rely on multiparty computation and cryptographic mechanisms, the proposed approach is applicable in the context of mobile networks, where devices have limited computation power and battery. The proposed solution in fact is based on symmetric encryption only, and reduces the impact of multiparty computation on the end to end communication and on power consumption. This comes however at the price of a slight increase in the bandwidth. Of course, there is also a clear trade-off between anonymity and latency overhead: the further we forward the packets, the better the anonymity is but the more is the latency overhead. To quantify that trade-off in practice, a prototype has been implemented that uses WiFi-enabled cellphones. The experimental results show that the overall impact on the end-to-end latency is negligible, thus ensuring applicability of the proposed solution to protect the privacy of real-time services (e.g., video streaming and voice activated services). In conclusion, there are a lot of open questions regarding the performance of the entire system, which should be carefully considered.

Business scenarios

The third line of investigation of this task considered the applicability of privacy metrics and privacy enforcing methods to real-world business scenarios. Three scenarios were considered:

- *Outsourcing.* Companies often have to outsource application development to third-parties. For application testing, they have to create samples of data which maintain the integrity of the application and, at the same time, do not disclose sensitive or private information contained in the production systems. Creating test data is often labor-intensive and needs to be tailored for specific contexts, resulting in long and expensive processes.

- *Support.* Technical support often needs data and the context (e.g., information on system configuration and parameters) to efficiently trouble shoot customers problems. Such information may be sensitive, and clients have not the will or the right to share them.

- *Analytics.* Analytics techniques are becoming more and more popular, but, at the same time, more complex. Businesses do not have always the skills to run them in-house and often rely on third-parties. Consequently, there is the need to disclose some data, without compromising security and compliance, but at the same time preserving the *relevant* properties of the dataset.

Privacy metrics and privacy enforcing methods (anonymization) provide the technical means for solving some of these issues. Most of these methods are still in the realm of research, and rarely used in real-world applications. In the first year of the project, we investigated the requirements emerging from these scenarios, and performed a gap-analysis to outline the main limitations of current approaches. In the second phase of the project, we will investigate, within the PrimeLife consortium, how novel approaches may be used to overcome these problems.

Quantifying the privacy level is very important for all the above mentioned scenarios, since a metrics will support the data holder in evaluating privacy risk and utility. Even if various measures have been proposed so far, they are still rarely used in real applications. As an example, we list a subset of issues.

- *Performance.* Most of the algorithms used for estimating privacy risk do not scale when huge amounts of real data are used. E.g., generating data for application testing may imply producing tables of several gigabytes, using probabilistic masking methods. For such applications current privacy metrics are still too time consuming.

- *Definition of quasi-identifiers.* Identifying which attributes, or combination of attributes, may be used for re-identification is sometimes a difficult task. E.g., context information used for troubleshooting contains system variables and it is often hard to define which of them may be used for re-identification.

- *Definition/access to dictionaries.* The basic idea in many privacy metrics definition, is trying to link the anonymized data to some external (not anonymized) data source (dictionary). Such dictionaries are often difficult to identify and access for the data holder before the data are actually released.

4.2.2 Techniques for enforcing data privacy (Task 2.3.2)

The work on the design of techniques for data privacy has followed two lines of investigation: *i)* analysis of requirements for the management of privacy in business applications; *ii)* management of privacy constraints in relational databases.

Requirements for management of privacy in business applications

The investigation identified a set of requirements for the above scenarios. Most of them may be addressed adapting existing anonymization methods, others need substantial improvements of the current methods (e.g., preserving semantics). Typical requirements include:

- *Preserving the datatype.* A basic requirement for using the data downstream. Numbers should be replaced with numbers, string with string.

- *Preserving syntax/format.* At an additional level of complexity, basic syntax should be conserved. It implies that the syntactical rule for each attribute must be considered. E.g., first vs. last digits in zip codes or credit card numbers.

- *Preserving semantics.* In some cases, there is the need to keep the 'meaning' of some attributes. Therefore, names should be replaced with meaningful names (possibly language-specific), diseases with diseases, ... To this scope, it may be needed, first, to have the necessary semantic information in the original dataset, then to have available databases with list of candidates for replacement.

- *Preserving Relationships.* Data themselves are often used as keys in relational databases. In particular, unique identifiers, as Social Security Number, may play this role. Accordingly the anonymization process should mask this data in a consistent way, to avoid to lose the relationships between tables.

- *Preserving the distribution of original data.* E.g., the percentage of empty fields, or the distribution of salaries in a salary table. This can be particularly relevant for heavy-tailed distributions, where extreme values have to be correctly sampled.
- *Preserving consistency.* Attributes are often correlated, so the anonymization process should be applied in a consistent way across multiple attributes. E.g., city, states, telephone numbers.

This investigation is the basis for the identification of techniques and tools able to support the construction of privacy-compliant business applications.

Privacy Constraints in Relational Databases

The work in this task focused on the design of techniques supporting the management of privacy constraints in relational databases. The motivation for this work derives from the consideration of data architectures that often emerge in the privacy scenarios, where a crucial issue is represented by the threat that derives from the integration of several sources of PII. Such scenarios range from traditional distributed database systems, where a centrally planned database design is then distributed to different locations; to federated systems, where independently developed databases are merged together; to dynamic coalitions and virtual communities, where independent parties may need to selectively share part of their knowledge towards the completion of common goals. Regardless of the specific scenario, a common point of such a merging and sharing process is that it is selective: if on one hand there is a need to share some data and cooperate, there is on the other hand an equally strong need to protect those data that, for various reasons, should not be disclosed.

The correct definition and management of protection requirements is therefore a crucial point for an effective collaboration and integration of heterogeneous large-scale distributed systems. The problem calls for a solution that must be expressive to capture the different data protection needs of the cooperating parties, as well as simple and coherent with current mechanisms for the management of distributed computations, to be seamlessly integrated in current systems and fully exploit the availability of technical solutions that are the fruit of a large amount of research and development.

This problem and a possible solution to it have been presented in a PrimeLife publication [DFJ$^+$08b]. The paper presents a novel and flexible representation of privacy constraints, expressed as access privileges (authorizations) to portions of a distributed relational schema. Authorizations regulate not only the data on which parties have explicit visibility, to ensure that query processing discloses only data whose release has been explicitly authorized, but also the visibility of possible associations such data convey. The proposed authorization form essentially corresponds to generic view patterns thus nicely meeting both expressiveness and simplicity requirements.

Data disclosure has been then captured by means of profiles associated with each data computation that describe the information carried by the relation. Given a set of authorizations and a query plan, it is then necessary to determine how the operations specified in the query have to be executed and who can execute them. For this purpose, the model is coupled with an algorithm that, given a query plan determines whether it can be safely executed and produces a safe query planning for it.

Intuitively, safe query planning means that the operations composing the query are executed by servers in such a way that the views (profiles) entailed by the operations are allowed by the authorizations. This analysis takes into account the way in which relational systems execute relational queries, with specific attention paid to the realization of the join operator. The main advantage of the proposed approach is its simplicity that, without impacting expressiveness, makes it nicely interoperable with current solutions for collaborative computations in distributed database systems. This is an important feature in distributed settings, where the minimization of data exchanges and the execution of steps of the queries in locations where it can be less costly, is a crucial factor in the identification of an execution strategy characterized by good performance.

4.2.3 Efficient organization and access to privacy-preserving data collections (Task 2.3.3)

The goal of this task is the investigation of techniques able to support efficient access to large collections of private information. The assumption is that the server storing the data has only partial access to the data, because the private information is completely or partially protected from access, typically using encryption. These techniques can find application in many contexts: the classical scenario is represented by data outsourcing, where the storage and dissemination of private resources relies on the services of a party that is not trusted with respect to the confidentiality of the data. Another scenario is represented by the use of encryption on specific sensitive data stored internally on a server which is trusted: the use of encryption offers protection against possible compromises of the server at the physical level; access to the encryption key occurs only when a legitimate access request requires the protected piece of information.

In all these applications, the support of large collections of data requires the definition of indexing structures, able to support the efficient execution of access requests interested only in a subset of the data. Indexing structures have been carefully studied in the past to support the execution of queries in traditional DBMSs; the consideration of the confidentiality requirements forces a redesign, as all the classical index types offer a clear opportunity for violating confidentiality by considering their structure. Also, techniques can be designed considering a variety of threats and typically a trade-off will be observed between the efficiency of the technique and the level of confidentiality against various threats it will be able to exhibit.

The work in the first year has considered techniques presented in the past and the corresponding threat models and has started to investigate a scenario with stronger threat models, taking specifically into account the attacks that a server registering the history of access requests is able to realize; we note that many of the techniques proposed in the past only consider a static configuration, where requests are separately considered. A few approaches have been proposed in the research community to protect against these attacks; unfortunately, the solutions offering the highest degree of protection (e.g., Private Information Retrieval) exhibit performance incompatible with what is required by real systems; on the other hand, solutions with a better performance profile (HMACbased indexes, or solutions recently proposed that assume of the availability of a secure processor module) are not able to provide protection against all the threats. The work in this task has considered recent approaches and techniques proposed in the past. The work has not yet produced scientific publication; the plan is to produce at least one scientific publication in Year 2 of the project.

4.3 Future research

4.3.1 Privacy assessment and privacy metrics (Task 2.3.1)

Considering the lines of investigation presented earlier, the plan for Task 2.3.1 is to continue the investigation mostly on the definition of privacy-aware models for hybrid networks and on the definition of privacy metrics for business applications.

In the context of privacy-aware solutions for hybrid networks, many interesting research directions will be analyzed. Among them, the enhancement of the communication algorithms between peers, mobile network operators and servers will be investigated, aiming at the development of an abstract design of a suitable privacy solution for hybrid mobile networks. Also, the consideration of a comprehensive threat model that includes untrusted mobile network operators and malicious and uncooperative peers will be considered, to evaluate the privacy solution in a more general and challenging scenario. Finally, a solution that provides automatic economic incentives for the neighbor peers to participate in the anonymizing network will be studied. All these aspects will be driven by two main requirements: *i)* the limitation of expensive communication or cryptographic operations (including the use of multiparty computation) to ensure applicability in a mobile environment, where low computation overhead and limited battery consumption are important goals; *ii)* the protection of the system against possible abuses of anonymity by providing operators with the ability of distinguish genuine vs malicious traffic.

With respect to privacy metrics for business applications, the next step will be trying to adapt privacy metrics to, at least, one of the scenarios presented before. The work in Task 2.3.1 will also investigate, with consortium partners, how novel metrics may be defined to address the above mentioned issues.

4.3.2 Techniques for enforcing data privacy (Task 2.3.2)

Both the research topics studied within Task 2.3.2 in the first year of the project will be extended.

The investigation on the management of privacy constraints in relational databases will consider richer definitions of constraints, extending the power of the logical language. An interesting potential lies in the definition of an approach that integrates in a single language the representation of privacy constraints and the visibility requirements set by the application. The goal would then be the definition of an approach able to build, in an automatic or semi-automatic way, a system configuration able to support both components of the model, possibly maximizing some performance metric on the cost required for the processing of a predefined query load on the system.

The research on techniques for privacy-conscious business applications will identify which methods are better adapted to the business scenarios considered in the first year of the project. As in the case of privacy metrics, particular attention will be devoted to performance. In addition, requirements will be further refined, taking into account some areas not addressed so far, such as usability and legal aspects.

4.3.3 Efficient organization and access to privacy-preserving data collections (Task 2.3.3)

The work in Task 2.3.3 will first aim to define novel techniques able to overcome some of the shortcomings of current proposals. The task will carefully consider the scenarios and the techniques developed in the other tasks and work packages, aiming to a solution able to work together with the other tools designed in the project.

The design of the technique will be described in papers that will be submitted to scientific conferences and journals. A prototype is also planned to demonstrate the behavior of the technique. Integration with the other prototypes will be carefully considered, due to the mutual benefits that can derive from the realization of a system demonstrating the cooperation among several advanced privacy services.

Chapter 5

Access control for the protection of user-generated data (WP2.4)

5.1 Introduction

The widespread access to information and communication channels provided by modern technology has introduced significant benefits allowing users to enjoy electronic services while releasing personal information needed for using the services. The increased power and interconnectivity of computer systems and the advances in memory sizes, disk storage capacity, and networking bandwidth allow this vast amount of personal information to be collected, stored, and analyzed in ways that were impossible in the past due to the restricted access to the data and the expensive processing (in both time and resources) of them. This situation has led to growing concerns about the privacy of their users, limiting the social acceptance of the electronic society.

The main goal of this work package is then to define new models and methods for the definition and enforcement of access control restrictions on user-generated data. These solutions should enhance the user awareness and empowerment, granting users the ability to participate in (and be aware of) the management and dissemination of their data and resources. This is a fundamental aspect for enabling users to live in an electronic society and to enjoy electronic services in the full respect of their privacy. We however observe that the specific problems that need to be addressed depend on the underlying assumptions. This work package therefore considers different scenarios where one major differentiating assumption is related to whether the privacy of the data is to be preserved also against the party receiving the data. Work package 2.4 has been then organized in the following three tasks, addressing different aspects of the privacy problem.

- *Task 2.4.1 Dissemination control and secondary use restrictions* addresses the problem of defining solutions for regulating the secondary use of personal information (i.e., subsequent uses other than the one for which information was initially released). The solutions will allow the specification and enforcement of constraints that data holders (i.e., the party storing and managing the data) should obey on users' data as they are transmitted and shared with other parties. Within this tasks, data holders are assumed to be trusted with respect to the enforcement of the secondary use restrictions.

- *Task 2.4.2 Access control to confidential data stored at external services* addresses the problem of enforcing access control to user-generated data that users wish to publish at an external honest-but-curious server, meaning that the server while trustworthy to properly manage the data, may not be trusted by the data owner to read their content.
- *Task 2.4.3 User-managed access control to personal data stored in trusted services* addresses the problem of defining solutions for the management of user-generated data stored on trusted servers, meaning that they act as expected with respect to the enforcement of policies.

These problems pose several new challenges to the design and implementation of technical solutions for the realization of access restrictions controlled by the user. For instance, with respect to the definition of secondary use restrictions, a negotiation process between a user and a data holder is needed to reach an agreement on the data handling restrictions. Another challenge relates to the fact that since servers may not be always under the control of trustworthy authorities, it is necessary to provide measures to protect the integrity and confidentiality of sensitive personal data and of the privacy control policies.

In the following, we first describe the advancement status of the research work done in the first year of PrimeLife. We then outline future work.

5.2 Research results

The research activity has regarded several thematics related to the development of solutions for enabling the enforcement of access restrictions on data generated from and/or by users. Such an activity resulted in four publications which appeared in international conferences and one book chapter.

5.2.1 Dissemination control and secondary use restrictions (Task 2.4.1)

The work in this task is started with an analysis of the requirements and current solutions for supporting user-controlled information dissemination. The task focuses on the protection of any type of personal identifiable information (PII), including location information for which the users should be able to explicitly identify how this information will be disclosed, when this can happen, and to whom this information will be revealed [ACDS08a]. In conjunction with Activities 5 and 6, we also studied how secondary use restrictions may be expressed in terms of policies, and how they may be enforced on the server side. To illustrate the typical issues that we may face in this task, we outline a scenario that may help in defining the relevant requirements and that can be seen as a test bed for further analysis. The scenario was agreed with partners in Activity 5 and Activity 6, we stress in this context the aspects related to secondary use of data. We considered the electronic Curriculum Vitae (eCV) scenario. A user creates an electronic CV (eCV) that contains up-to-date information on his personal details, work experience, academic qualifications, and a reference. The personal information could be entered by the user or provided by an official authority service. The other types of information services that certify what the user claims could be a university, recommendation and previous or existing employer. Each contributory data provider could have a rule or sticky policy attached to the data that outlines how the data will have to be handled when used by the data producer, data consumer, or third party. The parts of the eCV that contain this information cannot be altered by the applicant in order to preserve the policy preferences of the different services. These constraints, imposed by such data providers, may restrict the exposure of some information which is related to the company that should not be revealed. The eCV may be then submitted to a job broker, upon comparison between sender/receiver privacy policies.

As a first step, we have been analyzing how such secondary use requirements may be expressed using existing policies languages, such as XACML, EPAL, and PRIME policy languages (the latter two, along with other Activity 5 consortium partners). We identified limitations in the current available methods, in particular the difficulties to compose different policies in a single one when conflicts occur. Particular attention was devoted to XACML due to its focus on access control, and we investigated how it may be extended to include more complex obligations and to compose different policies.

Furthermore, we are investigating the problematic area of policy enforcement which normally relies solely on static trust measures or a service level agreement between different parties. In the context of access control and Service Oriented Architecture (SOA), we are focusing more on providing measures that allow for dynamic enforcement in a loosely coupled environment.

We also investigated current techniques used in commercial systems for guaranteeing access control on user-generated data, and providing transparent measures to facilitate auditing.

5.2.2 Access control to confidential data stored at external services (Task 2.4.2)

The work in this task has been focused on the definition of techniques and models for protecting the personal information of users when it is stored on external servers. Since the techniques that should be in place depend on the kind of information and how it is stored, we consider a reference scenario where information can be possibly encrypted and can be related to any personal aspect of the users (e.g., healthcare, financial, and biometric information). Within this task, two main contributions have been provided: *i)* an approach for protecting the confidentiality of the privacy policies, and *ii)* an approach for protecting personal information derived from human biometric traits.

Protection of the confidentiality of privacy policies

The first contribution is strictly related to the general problem of protecting personal information when stored on external “honest-but-curious” servers, which has recently received considerable attention by the research community and for which some advancements have been proposed [HIM03]. In this context, selective encryption techniques have been proposed for safely delegating to the external server itself also the management of the access control policy while ensuring complete protection of the content as well as correct enforcement of the policy against possible misbehaviour and collusions [DFJ$^+$07a, DFJ$^+$07b]. The basic idea of these solutions is to combine authorization-based access control and cryptographic protection by defining an encryption policy equivalent to the authorization policy to be enforced on the resources. The encryption policy exploits a *key derivation method* [AT83, AFB05, CMW06, San87], which operates on a key derivation graph computing the keys of lower-level vertices based on the keys of their predecessors and on information publicly available (tokens). Here, the graph includes a vertex v for each possible set of users and an edge (v_i,v_j) for all pairs of vertices such that the set of users represented by v_i is a subset of the set of users represented by v_j. The encryption policy is then created in such a way that: *i)* each resource is encrypted with the key of the vertex representing its access control list, and *ii)* each user knows the key associated with the vertex representing the user in the graph.

Along this line of research, an interesting problem that has been investigated in this task is the protection of the confidentiality of the access control policy enforced by the external server. Indeed, the use of a key derivation graph and its tokens, while greatly simplifying key management, introduces however a new vulnerability related to policy confidentiality. As a matter of fact, public availability of tokens, and therefore of the corresponding key derivation graph, makes visible the relationship between users and resources they are authorized to access, and therefore discloses the authorization policy. In several contexts, however, the policy itself should be considered confidential as owners do not wish to publicly declare to whom they give (or not give) access to their resources. Also, an analysis of the policy may allow observers to reconstruct the structure of the social network of users accessing the system and of their real identities. Since the overall aim of these novel solutions is to allow an efficient confidentiality-preserving mechanism for resource dissemination, the protection of the access control policy appears a natural requirement that systems will be interested in supporting, as long as system performance remains guaranteed.

The confidentiality problem of the access control policy and a possible solution to it have been presented in a publication [DFJ$^+$08a]. The proposed solution basically provides a protection (encryption) layer to the policy information exploiting the key derivation graph itself. The idea is to protect the topology of the key derivation graph by encrypting the tokens in such a way to preserve the ability of each user to retrieve the tokens needed to derive the keys that should be used to decrypt the resources she can access. To this purpose, given a token $t_{i,j}$ that allows the derivation of the key associated with vertex v_j starting from the key of vertex v_i, the token is encrypted by using the key associated with vertex v_i. In this way, we have the guarantee that the decryption operation can be performed only by users that directly or indirectly are authorized to know the key associated with v_i and therefore that are also authorized to know the key associated with vertex v_j. The main drawback of this approach is that the process for deriving a specific key becomes expensive since, in the worst case, to derive the key associated with a particular vertex a user can only perform a top down traversal of the key derivation graph, starting from the vertex corresponding to her key. The user therefore interacts with the server to progressively retrieve the tokens that allow the derivation of the key associated with her descendant vertices, until she reaches the key of interest or the visit terminates. The fundamental observation for solving this drawback is that to check whether a given user whose key is associated with vertex v can derive the key associated with vertex v_j , it is necessary to check whether vertex v_j is reachable from vertex v through a path in the graph. We then propose an algorithm for computing the transitive closure of the ancestor-descendant relationship. The transitive closure is represented through numerical intervals associated with the vertices in the graph. By looking at the intervals associated with a specific vertex v_i, it is then immediate to verify whether there is a path from v_i to another vertex v_j. To describe the performance of the technique proposed, its performance has been analyzed along three directions. The first analyzes the performance in terms of the number of vertices visited on the key derivation graph compared to the number of vertices visited by a blind search that does not use the information (intervals) about the transitive closure of the ancestor-descendant relationship. The second estimates the average number of vertices that have to be visited to reach, starting from a user vertex, all the vertices in the key derivation graph. The third evaluates the storage requirements necessary to materialize the whole transitive closure. The experiments performed confirm that the proposed solution is manageable, even for large graphs and scenarios with large user populations and complex protection requirements.

Protecting personal data derived from biometric traits

The second contribution is related to the protection of a particular type of personal information: biometric traits. Until a few years ago, the use of biometric systems was almost limited to security needs. We now assist to an increasing interest for biometrics in the research community and many biometric commercial products are also becoming available. However, side to side with the widespread diffusion of biometrics an opposition grows towards the acceptance of the technology itself. Two main reasons might motivate such resistance: the *reliability* of a biometric system and the possible threats to users' *privacy*. In fact, a fault in a biometric system, due to a poor implementation or to an overestimation of its accuracy could lead to a security breach. The protection of biometric traits is then particularly challenging since they cannot be strictly considered as "secrets"; they can be inadvertently disclosed (e.g., fingerprints are left on a myriad of objects such as doors' handles or elevator buttons; pictures of faces are easily obtained without the cooperation of the subjects). Moreover, if biometric traits are captured or if their digital representations are stolen, they cannot be simply replaced or modified in any way, as it can be done, for example, with passwords [Sch99]. For this reason, privacy agencies of many countries have ruled in favor of a legislation which limits the biometric information that can be centrally stored or carried on a personal ID. For instance, templates (i.e., mathematical information derived from a fingerprint) are retained instead of the picture of the fingerprint itself. Also, un-encrypted biometrics are discouraged.

These aspects have limited so far the number of applications in which biometric authentication procedures were allowed by privacy agencies in several countries. In addition to this, users often perceive the potential threat to their privacy and this reduces the user acceptance of biometric systems, especially on a large scale. To avoid these problems, a privacy-aware biometric cryptographic scheme has been presented in two publications [CGP+08a, CGP+08b]. Such a schema enables the creation of a unique identifier associated with each enrolled person by exploiting the error tolerant properties of the biometric templates. This is obtained by using multiple biometric traits concurrently and the recently introduced cryptographic primitives secure sketches and fuzzy extractors. The resulting scheme is multimodal, in the sense that multiple biometric traits (at least two) can be used. The proposed system is composed of two main modules: the enrollment module and the verification module. The enrollment module creates an ID starting from the biometric readings of a user. The ID can be envisioned as a function of the biometric traits and is associated with the owner of the biometric traits. The ID is then stored or printed on a document and must be provided during the verification phase. The verification module verifies the identity claimed by the user using the ID and novel biometric readings. The process is successful if the novel readings match the ones used to build the ID.

The basic enrollment and verification modules can also be combined hierarchically and/or in parallel (with respect to the input biometric readings) to implement authentication applications having different levels of security and using a higher number of biometric features. The hierarchically composition exploits multiple enrollment modules that are applied in cascade. To better illustrate the idea, consider a two-level hierarchical composition and suppose that each enrollment module receives two biometric traits as input. In this case, three biometric traits are needed: two of them are the input of the first enrollment module, whose output together with the third biometric trait represents the input of the second enrollment module. The parallel composition offers a simple method to exploit different biometric traits to create the ID. The idea is that each input of the enrollment module is obtained by combining more biometric traits. Consequently, the level of multi-modality implemented is higher than in the basic approach since more than two biometric traits are in use. To verify the applicability of the proposed model, it has been implemented by taking into consideration two biometric traits: iris and fingerprint. The implementation shows the feasibility of the scheme and offers an idea of the performances one might obtain from the application on real datasets. Indeed, the resulting error rate is acceptable and it is not worse than the best error rate of the single-trait biometric systems on which it is based. The work paves the way for large scale applicability of privacy-aware biometric systems.

Related to the development of a privacy-aware biometric cryptographic system, there is the problem of how the acquisition of the biometric traits is performed. An interesting approach consists in using low-cost sensors, which is becoming a significative possibility especially if we consider fingerprint images. In fact, the possibility to obtain fingerprint images by using lowcost cameras and off-the-shelf webcams has a great practical importance since nowadays fingerprint biometrics can be successfully achieved only by using dedicated sensors. However, the images of human fingerprints acquired from these kinds of sensors are very different from the images obtained by dedicated fingerprint sensors. To exploit the paramount variety of techniques now available to achieve verification and recognition by fingerprint biometrics, it is then important to process the camera image of the fingerprints so that they are as similar as possible to the images produced by dedicated sensors. To address these new issues, an approach that allows for exploiting a webcam as a fingerprint biometric sensor on a personal computer has been presented in a publication [PS08b]. The proposed technique encompasses a prefiltering step, the segmentation of the fingertip image, a fingertip registration phase, the dedicated processing techniques for the ridge enhancement, and a post-processing phase. The efficacy and efficiency of the proposed technique has been evaluated by comparing it with the use of commercial dedicated sensors. The results are incouraging and show that the proposed approach is feasible and that the behavior of the low-cost system is comparable in accuracy to the system exploiting the dedicated sensor.

5.2.3 User-managed access control to personal data stored in trusted services (Task 2.4.3)

The work in this task has investigated dynamic access control mechanisms that can be driven by a user to give access to their personal data as needed for processing by other services (e.g., in a service composition).[1] The main differences with Tasks 2.4.1 and 2.4.2 is that in this task there is the assumption that the service storing the personal data is trusted by the user, meaning that the server will handle personal data according to the data handling policy associated with the data themselves.

Figure 5 illustrates three possible deployments of personal data and required mechanisms to let user control his personal data. Each option is described in next paragraphs.

Deployment 1: Data store controlled by data subject (Figure 5(a)). The data subject (user) runs the data store (e.g., home server) or is registered to an external data store (e.g., cloud service). In both cases, the user trusts the data store that can specify access control policies (i.e., define who can access a given personal data).

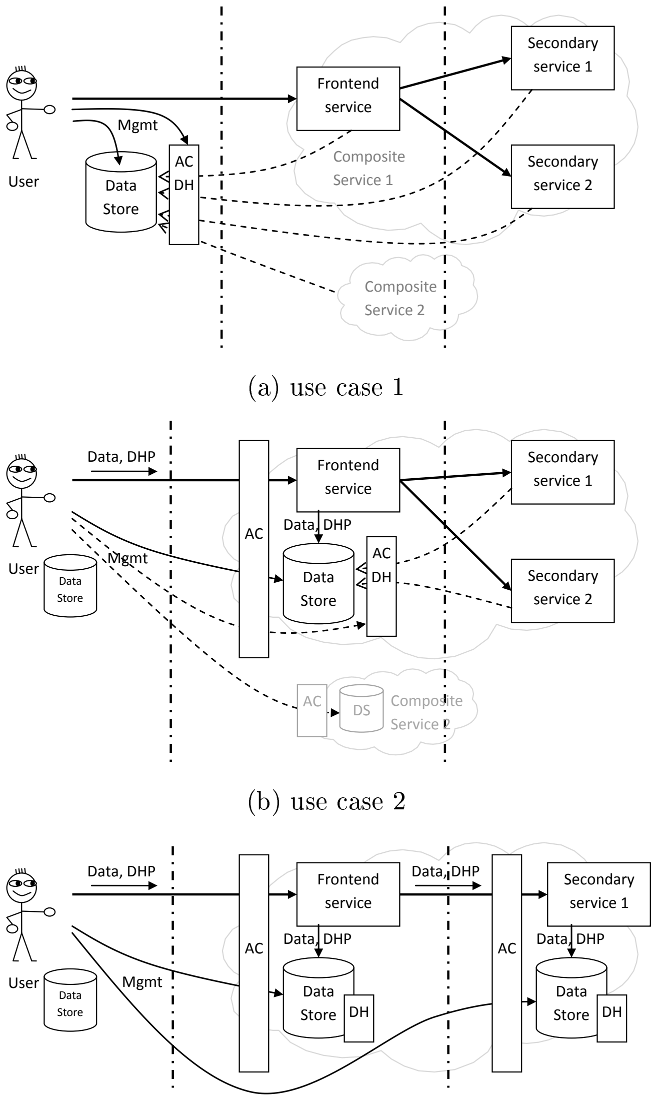
Any service requiring personal information gets a pointer on the data and a credential to authenticate. Services are required to request data for each “transaction”. Key features of this deployment are:

- *User control*: personal data are not distributed and the user can know who has access to given personal data, the user can update the data at any time, the user can deny access at any time, the user can log access to her personal data.
- *Service composition*: each service must have a way to get the data. No guarantee that data will be available (e.g., data store can be shut down, access may be denied, access may trigger a long-lasting approval process).

For instance, this approach is already in place with online health records such as HealthVault (http://www.healthvault.com/) or Google Health (https://www.google.com/health).

Deployment 2: Data store controlled by data processor (Figure 5(b)). The user provides personal data to a data processor (front-end service of composition). The front-end service stores personal data and offers a management interface to the user. Any secondary

[1] Note that this work will reuse the data handling policy (DHP) language currently under work in Activity 5. Current work in WP5.2 regarding the links between access control policies and DHP is especially relevant for this task.



(a) use case 1

(b) use case 2

(c) use case 3

Figure 5: Three possible deployments of personal data

service, i.e. any third party that gets the personal data from this front-end service, must authenticate to the front-end service. Key features of this deployment are:

- *User control*: personal data are distributed to well define services. The user must have a way to manage her data and thus should have a credential to authenticate to the data processor.
- *Service composition*: within the service composition, data are available. Since data are stored and reused, it is necessary to keep track of data handling policies

	use case 1	**use case 2**	**use case 3**
User control	Full	Medium	Complex
Service composition			
- Data access	Difficult	Medium	Easy
- Data handling	Simple	Complex (front-end)	Complex (all services)

Table 3: Comparing three use cases

associated with the data.

This approach is somehow in place in some workflows. For instance, a hospital workflow could centralize patient data and share them with different services or even external entities (e.g., another hospital or insurance company).

Deployment 3: Personal data is heavily distributed (Figure 5(c)). The user provides personal data to data processor (front-end service of composition). This front-end service stores personal data and potentially shares personal data with other services of the composition (secondary services), which also store it. Key features of this deployment are:

- *User control*: personal data are distributed to hidden services (e.g., hotel booking service behind front-end travel booking service). The user must have a way to manage her data and thus should know where her data were distributed. Moreover, the user should be able to authenticate to each management interface.
- *Service composition*: data are available to each service. Each service must properly deal with data handling policies associated with stored data.

Nowadays, this is the most usual deployment. When privacy is taken into account, data handling policy is properly taken into account. However, users generally have no control on their data, that is, they cannot easily update them.

Table 3 describes the complexity of the three deployments above-mentioned from a user and a server perspective. Depending on the type of service, all cases can be found in deployed services with more or less support for privacy. Task 2.4.3 will mainly study shortcomings of such existing deployments and propose solutions to enable user control (see Section 5.3.3).

5.3 Future research

Issues to be investigated include the following.

5.3.1 Dissemination control and secondary use restrictions (Task 2.4.1)

It is planned to deepen our understanding in the research areas mentioned in Section 5.2.1, particularly access control and enforcement, with the aim to synchronize with the work of Activities 5 and 6. This work will contribute to a prototype that will include a policy engine in Activity 5.

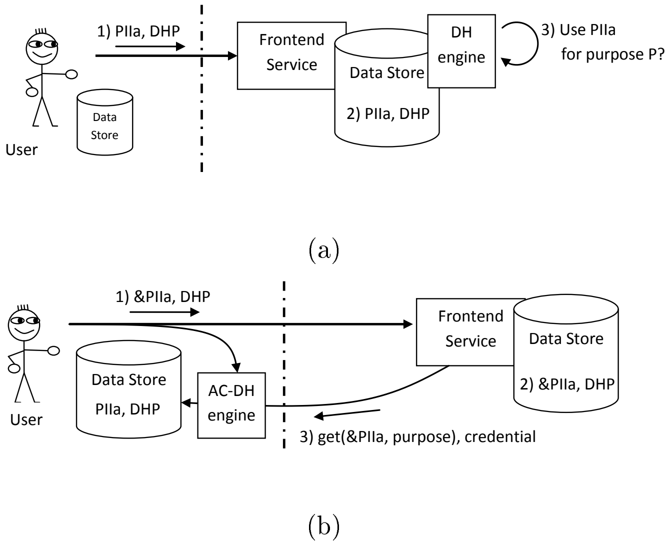
5.3.2 Access control to confidential data stored at external services (Task 2.4.2)

Within the reference scenario there are different open issues that we plan to address in the future research.

A first crucial problem to be addressed concerns the definition of a formal model for representing and transforming an authorization policy through an equivalent encryption policy. Such a model will be developed by taking into consideration an important requirement related to the number of keys and tokens needed to correctly represent an authorization policy via an encryption policy. In fact, the management of a huge number of keys and tokens can be unfeasible in practice. The goal will be then the definition of minimal encryption policy and the development of an algorithm for computing a minimal encryption policy equivalent to a given authorization policy. Since the minimization of the number of keys and tokens maintained by the system and distributed to users is a NP-hard problem, we plan to define different heuristic approaches for its resolution. We will then experimentally evaluate the performance of the proposed algorithms, possibly comparing them with previous approaches.

Since the access control policy is likely to change over time, another crucial problem is the definition of an approach for supporting policy updates in dynamic scenarios. We then plan to define an approach for supporting policy changes when the data owner needs to maintain the control on the authorization policy management. We will therefore illustrate how the key derivation graph can be modified in case of a dynamic scenario, where the access control policy may change. Intuitively, in this case, the resource whose access control list is changed, has to be downloaded from the server, decrypted through the old key, re-encrypted through the new key, and the new encrypted version of the resource has to be uploaded on the server. However, in scenarios involving potentially huge sets of resources of considerable size, re-encryption and re-transmission by the owner may not be acceptable. Therefore, building on the formal model, we plan to define a two-layer approach to outsource, besides the resource storage and dissemination, the authorization policy management (the advantage compared with a solution requiring to re-send a novel encrypted version of the resource is typically huge and arbitrarily large). We will evaluate the robustness of the approach against attacks from users who access and store all information offered by the server, or against collusion attacks, where different users (or a user and the server) combine their knowledge to access resources they would not otherwise be able to access. An important strength of our solution will be that it will not substitute the current proposals, rather it will complement them, enabling them to support selective encryption and easily enforce dynamic policy changes.

With respect to the protection of biometric traits, we plan to study the performances on the proposed techniques on larger datasets and also in varying operational conditions. Also, we plan to formally model the security properties of the proposed privacy-aware biometric cryptographic scheme.



(a)

(b)

Figure 6: User sending PII to a Data Processor (a) and user sending a pointer to PII (b)

5.3.3 User-managed access control to personal data stored in trusted services (Task 2.4.3)

Current authorization languages such as SecPAL [BFG07, Dil06] and XACML [Mos05] combined with current authorization protocols such as WS-Trust [TC07] and former work on SeCSE delegation framework [BNN$^+$08, BN08], solve the major authorization requirements of use case 1 and part of authorization requirements of use cases 2 and 3 (see Section 5.2.3). The key remaining issue is to define obligations related to user control. In the following, we analyze three different issues that will be addressed as future work.

Replacing personal data by a reference (use cases 1 and 2)

To replace data handling by access control, the data subject must provide to the data processor a reference to the personal data and grant access to this one. Figure 6(a) shows the usual behavior, where the data subjects provide personal data to the data processor. For the sake of readability, we call $\&x$ the address of x or a pointer to x. For instance:

```
&PIIa = {service:http://www.contoso.com/serviceX;
         action:getData;
         parameter: 12345678-1234-5678-1234-567812345678;
         credential: AuthZ Credential XYZ }
```

Figure 6(b) shows the data subject that provides a references to her personal data $\&PII_a$ (step 1) that is then stored (step 2). Next, the data processor retrieves the data when required by using the reference including a credential to authenticate to the data

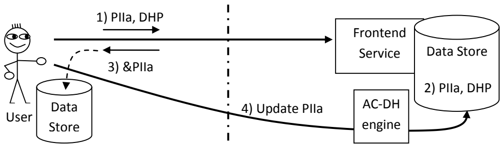


Figure 7: Obligation to provide management rights

store (step 3). Ideally, the fact that personal data are not stored on data processor side should not impact the application.

Different protocols could easily be supported to retrieve the data, (e.g., SOAP and WS-Security) as well as different types of credentials could be used (e.g., SecPAL, anonymous credentials [Bra00, CL01b], or encrypted user-name and password). When the front-end service requires sharing collected personal data with third parties (secondary services), delegation mechanisms become relevant. When the front-end service needs to use the personal data (e.g., PIIa) for a specific purpose, it must contact the data subject that takes an access control decision.

Future work will be then devoted to investigate the state of the art regarding Unique Identifiers (UId) and most suitable mechanisms to associate rights (i.e. credentials). Existing access control mechanisms will be used or enhanced to take into account considerations specific to data-handling (e.g., purpose).

Obligation to provide management rights (use cases 2 and 3)

There is no universal list of obligations. First, some policy languages (e.g., EPAL [AHK$^+$03], XACML [Mos05], and PRIME-DHP [ACDS08b]) offer a placeholder for obligations but do not specify them. Second, current research proposals on obligations [HBP05, MB07] define a framework to enforce obligations but do not provide an exhaustive list of obligations. Finally, apart a small set of usual obligations (e.g., delete PII, notify data subject, log PII usage), most obligations are domain specific.

Not surprisingly, an obligation to let data subject modify her personal data is not documented in the literature. One key challenge is to authenticate the user when using such a management interface. It does not seem suitable to create a dedicated account/credential that must be stored by the user. As long as the data subject uses the same credential (without unlinkability features) to provide personal information and to manage it, the data processor may authorize the data subject and may propose management features.

Figure 7 illustrates the data subject (user) that provides some personal data (PIIa) to a data processor. A data handling policy (DHP) is attached to the personal data (step 1). The DHP contains a management obligation stating that the data consumer must let the data subject update or remove her data at any time. The data subject gets a reference to the data (&PIIA), including authorization and/or authentication credentials (step 3). Finally, the data subject authenticates to the data processor to manage personal data (step 4).

Future work will be then devoted to specify such an obligation scheme. We will define protocols to commit to such an obligation as well as specifying parameters (e.g., address of the management endpoint, credential). We will define the structure of the obligation and mechanisms to propagate such obligations when data are shared with third parties (secondary services).

Callback obligation and associated authorization (use cases 2 and 3)

Callback obligations (i.e., data subject notification) are often mentioned [HBP05]. However, they generally neither take authentication nor authorization into account. This callback obligation should contain a credential to authenticate the data subject. This may be a dedicated credential (e.g., SecPAL) or a reference to an existing credential (e.g., X.509 certificate of the data processor). When personal data are distributed to multiple parties, any service has to fulfill the obligation. Some form of delegation may be necessary (e.g., deployment 3) or the front-end service may act as a proxy forwarding callbacks (e.g., deployment 2).

It is important to make sure that incoherent obligations cannot be specified. Indeed, the fact that a callback obligation is attached to personal data should be sufficient to convince the data subject that the data processor is authorized to fulfill the obligation. In other words, the obligation in the data handling policy attached to the personal data must be in sync with the access control policy of the data subject. Note that when the data subject authenticates with anonymous credentials, anonymous callback mechanisms would be also required.

Future work will be then devoted to specify such an obligation scheme. Protocol to commit to such an obligation as well as defining parameters (e.g., callback address, credential) will be studied. The structure of the obligation will also be documented. Mechanisms to propagate such obligation when data is shared with third parties (secondary services) will be addressed as well. This may require some form of delegation of rights. Finally, a research demonstrator implementing and showcasing deployments 1, 2, and/or 3 will be provided. The focus will be on the two types of obligations mentioned above.[2] The usability and security of the three types of deployments will be briefly documented from user and composition points of view.

[2] Note that results of Activity 5 (Work Package 5.3) will be reused to cover data handling and access control. Results from work package 6.3 may be reused to tackle composition of different services with such obligations.

Chapter 6

Conclusions

During the first year of PrimeLife, the work packages of Activity 2 have made, and continue to make, significant contributions to the development of privacy-aware techniques for ensuring privacy and trust in the electronic society. WP2.1 has improved the state-of-the-art in different cryptographic areas related to anonymity and to the protection of the users privacy in general. WP2.2 has worked towards the analysis of: transparency support tools; privacy measurements for evaluating the degree of privacy provided to the users; and privacy aspects regarding collaborative groups and reputation systems. WP2.3 has developed new solutions for enforcing privacy constraints in mobile networks and in real-world business scenarios, and has analyzed the current privacy metrics for data collections. WP2.4 has developed novel solutions for efficiently protecting the confidentiality of privacy policies and for protecting sensitive information derived from human biometric traits. WP2.4 has also provided a first analysis of how to express secondary use restrictions and has investigated dynamic access control mechanisms that can be driven by users for providing access to their data.

The reported research results have been published in different leading international conferences, including ACM CCS, the flagship conference of the ACM SIGSAC group, IEEE ICDCS, ASIACRYPT, and ACSAC.

The research results presented in this report represent only a first step towards the realization of the overall goal of PrimeLife. The research will continue towards the development of novel and innovative solutions for ensuring privacy of the users in the electronic society. Open issues that will be investigated include: the development of solutions for an efficient revocation of anonymous credentials; the development of mechanisms for supporting privacy aware third-party services; advancements in the areas of biometric authentication, trust wallet, and private service access; the development of a generic transparency tool; the development of techniques for supporting privacy-aware collaborative groups and reputation systems; the enhancement of user awareness; the definition of privacy metrics for business applications; the enhancement of the privacy-aware communication protocols in hybrid networks; the investigation of richer definitions of privacy constraints in relational databases; the definition of a formal model for transforming an authorization policy into an equivalent encryption policy as well as the development of a solution supporting policy changes in a scenario where the information is stored at an external server; the performance evaluation of the techniques developed for protecting biometric traits; and the definition of an obligation schema.

Chapter 7

Abstracts of research papers

7.1 Cryptographic mechanisms (WP2.1)

1. J. Camenisch, M. Kohlweiss, and C. Soriente, “An Accumulator Based on Bilinear Maps and Efficient Revocation for Anonymous Credentials,” in *Public Key Cryptography, 12th International Workshop on Practice and Theory in Public Key Cryptosystems (PKC 2009)* [CKS09].
   Abstract. The success of electronic authentication systems, be it e- ID card systems or Internet authentication systems such as CardSpace, highly depends on the provided level of user-privacy. Thereby, an important requirement is an efficient means for revocation of the authentication credentials. In this paper we consider the problem of revocation for certificate-based privacy-protecting authentication systems. To date, the most efficient solutions for revocation for such systems are based on cryptographic accumulators. Here, an accumulate of all currently valid certificates is published regularly and each user holds a witness enabling her to prove the validity of her (anonymous) credential while retaining anonymity. Unfortunately, the users’ witnesses must be updated at least each time a credential is revoked. For the know solutions, these updates are computationally very expensive for users and/or certificate issuers which is very problematic as revocation is a frequent event as practice shows.
   In this paper, we propose a new dynamic accumulator scheme based on bilinear maps and show how to apply it to the problem of revocation of anonymous credentials. In the resulting scheme, proving a credential’s validity and updating witnesses both come at (virtually) no cost for credential owners and verifiers. In particular, updating a witness requires the issuer to do only one multiplication per addition or revocation of a credential and can also be delegated to untrusted entities from which a user could just retrieve the updated witness. We believe that thereby we provide the first authentication system offering privacy protection suitable for implementation with electronic tokens such as eID cards or drivers’ licenses.

2. J. Camenisch, M. Kohlweiss, A. Rial, and C. Sheedy, "Blind and Anonymous Identity-Based Encryption and Authorised Private Searches on Public Key Encrypted Data," in *Public Key Cryptography, 12th International Workshop on Practice and Theory in Public Key Cryptosystems (PKC 2009)* [CKRS09].
   Abstract. Searchable encryption schemes provide an important mechanism to cryptographically protect data while keeping it available to be searched and accessed. In a common approach for their construction, the encrypting entity chooses one or several keywords that describe the content of each encrypted record of data. To perform a search, a user obtains a trapdoor for a keyword of her interest and uses this trapdoor to find all the data described by this keyword.

   We present a searchable encryption scheme that allows users to privately search by keywords on encrypted data in a public key setting and decrypt the search results. To this end, we define and implement two primitives: public key encryption with oblivious keyword search (PEOKS) and committed blind anonymous identity-based encryption (IBE). PEOKS is an extension of public key encryption with keyword search (PEKS) in which users can obtain trapdoors from the secret key holder without revealing the keywords. Furthermore, we define committed blind trapdoor extraction, which facilitates the definition of authorisation policies to describe which trapdoor a particular user can request. We construct a PEOKS scheme by using our other primitive, which we believe to be the first blind and anonymous IBE scheme. We apply our PEOKS scheme to build a public key encrypted database that permits authorised private searches, i.e., neither the keywords nor the search results are revealed.

3. J. Camenisch, T. Groß, T.S. Heydt-Benjamin, "Rethinking accountable privacy supporting services: extended abstract," in *Digital Identity Management 2008* [CGHB08].
   Abstract. As privacy concerns among consumers rise, service providers will increasingly want to provide services that support privacy enhancing technologies. At the same time, providers of commercial services require the security of identifying misbehaving users. For instance, users that do not pay their bill can be held accountable for their behavior. We propose a scheme that permits privacy support while retaining accountability. In our proposed scheme an honest user may enjoy full anonymity, but dishonest users who do not pay their bill have their identity revealed. In contrast to existing revocable anonymity systems, our proposed scheme requires less trust in an external authority, while simultaneously making accountability easier (and less costly) to achieve. We contribute the concept of a time capsule, that is, a verifiable encryption with timed and revocable decryptability.

4. J. Camenisch, T. Groß, "Efficient attributes for anonymous credentials;" in *ACM Conference on Computer and Communications Security 2008* [CG08].
   Abstract. We extend the Camenisch-Lysyanskaya anonymous credential system such that selective disclosure of attributes becomes highly efficient. The resulting system significantly improves upon existing approaches, which suffer from a linear complexity in the total number of attributes. This limitation makes them unfit for many practical applications, such as electronic identity cards. Our system can

incorporate an arbitrary number of binary and finite-set attributes without significant performance impact. Our approach folds all such attributes in a single attribute base and, thus, boosts the efficiency of all proofs of possession. The core idea is to encode discrete binary and finite-set attribute values as prime numbers. We use the divisibility property for efficient proofs of their presence or absence. We additionally contribute efficient methods for conjunctions and disjunctions. The system builds on the Strong-RSA assumption alone. We demonstrate the applicability and performance improvements of our method in realistic application scenarios, such as, electronic identity cards and complex/structured credentials. Our method has crucial advantages in devices with restricted computational capabilities, such as smartcards and cell phones.

5. J. Camenisch, R. Chaabouni, A. Shelat, "Efficient Protocols for Set Membership and Range Proofs," in *Proc. of the ASIACRYPT 2008* [CCS08].
Abstract. We consider the following problem: Given a commitment to a value σ, prove in zero-knowledge that σ, belongs to some discrete set Φ . The set Φ can perhaps be a list of cities or clubs; often Φ can be a numerical range such as [1,220]. This problem arises in e-cash systems, anonymous credential systems, and various other practical uses of zero-knowledge protocols. When using commitment schemes relying on RSA-like assumptions, there are solutions to this problem which require only a constant number of RSA-group elements to be exchanged between the prover and verifier [5, 15, 16]. However, for many commitment schemes based on bilinear group assumptions, these techniques do not work, and the best known protocols require $O(k)$ group elements to be exchanged where k is a security parameter. In this paper, we present two new approaches to building set-membership proofs. The first is based on bilinear group assumptions. When applied to the case where Φ is a range of integers, our protocols require $O(\frac{k}{\log k - \log \log k})$ group elements to be exchanged. Not only is this result asymptotically better, but the constants are small enough to provide significant improvements even for small ranges. Indeed, for a discrete logarithm based setting, our new protocol is an order of magnitude more efficient than previously known ones. We also discuss alternative implementations of our membership proof based on the strong RSA assumption. Depending on the application, e.g., when Φ is a published set of values such a frequent flyer clubs, cities, or other ad hoc collections, these alternative also outperform prior solutions.

6. M. Belenkiy, M. Chase, M. Kohlweiss, A. Lysyanskaya, "P-signatures and Noninteractive Anonymous Credentials," in *Theory of Cryptography Conference (TCC 2008)* [BCKL08].
Abstract. In this paper, we introduce P-signatures. A P-signature scheme consists of a signature scheme, a commitment scheme, and (1) an interactive protocol for obtaining a signature on a committed value; (2) a non-interactive proof system for proving that the contents of a commitment has been signed; (3) a non-interactive proof system for proving that a pair of commitments are commitments to the same value. We give a definition of security for P-signatures and show how they can be realized under appropriate assumptions about groups with a bilinear map. We make extensive use of the powerful suite of non-interactive proof techniques due to Groth and Sahai. Our P-signatures enable, for the first time, the design of a

practical non-interactive anonymous credential system whose security does not rely on the random oracle model. In addition, they may serve as a useful building block for other privacy-preserving authentication mechanisms.

7. B. Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno, R. Mueller, "Evaluation Methodology for Analyzing Environment Influence in Biometrics," in *Proc. of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV)* [FSSRAMM08].
   Abstract. Biometrics is a new technology that provides a secure identification, which is being used more and more these days. However, sensors and biometric systems are usually only tested by their own suppliers, and rarely an independent evaluation of these products is carried out. Also, when evaluating a biometric system/device, many factors that can affect its performance, are not even considered. This is the case of the environmental conditions, i.e. humidity, temperature, etc. In this paper, authors propose an evaluation methodology for measuring environmental factors influence in biometric products and their applications. This methodology describes factors, tools, requirements and procedures, needed to perform this kind of performance evaluations.

7.2 Mechanisms supporting users' privacy and trust (WP2.2)

1. M. Bergmann, "Testing privacy awareness," in *Proc. of the IFIP/FIDIS Internet Security and Privacy Summer School* [Ber09].
   Abstract In web-based business processes the disclosure of personal data by the user is an essential part and mandatory for the processes. Privacy policies help to inform the user about his/her rights and to protect the user's privacy. In this paper we present a test to empirically measure how the user's privacy awareness changes by presenting specific elements of the privacy policy in close proximity to the required data items. We compare an experimental group using an enhanced interface to a control group using a conventional interface regarding their capability to recall the agreed privacy-related facts. A concrete online survey has been performed. The major results are presented.

2. S. Pötzsch, "Privacy awareness - a means to solve the privacy paradox?," in *Proc. of the IFIP/FIDIS Internet Security and Privacy Summer School* [PÖ9].
   Abstract People have limited memory capabilities, cannot keep in mind large amounts of information, cannot pay attention to too many things at the same time, and forget much information after a while; computers do not suffer from these limitations. Thus, revealing personal data in electronic communication environments and being completely unaware of privacy might nowadays cause a lot of privacy issues later. Even if users are privacy aware in general, the so-called privacy paradox shows that they do not behave according to their stated attitudes. This paper discusses explanations for the existing dichotomy between users' intentions towards disclosure of personal data and their behaviour and presents requirements on tools for privacy-awareness support in order to counteract the privacy paradox.

3. F. Pingel and S. Steinbrecher, "Multilateral secure crosscommunity reputation systems," in *Proc. of of the Fifth International Conference on Trust and Privacy in Digital Business* [PS08a].
   Abstract The Internet gives people various possibilities to interact with each other. Many interactions need trust that interactors behave in a way one expects them to do. If people are able to build reputation about their past behaviour this might help others to estimate their future behaviour. Reputation systems were designed to store and manage these reputations in a technically efficient way. Most reputation systems were designed for the use in single Internet communities although there are similarities between communities. In this paper we present a multilateral secure reputation system that allows to collect and use reputation in a set of communities interoperable with the reputation system. We implemented our system for the community software phpBB[1].

4. S. Steinbrecher, "Enhancing multilateral security in and by reputation systems," in *Fourth FIDIS International Summer School* [Ste08].
   Abstract With the increasing possibilities for interaction between Internet users exceeding pure communication, in multilateral security the research question arises to rethink and extend classical security requirements. Reputation systems are a possible solution to assist new security requirements. But naturally also reputation systems have to be designed in a multilateral secure way. In this paper we discuss both multilateral security by and in reputation systems. An overview on the possibilities how such systems could be realised is given.

5. S. Steinbrecher, S. Groß, Markus Meichau, "Jason: A scalable reputation system for the semantic web," in *Proc. of IFIP International Information Security Conference* [SGM09].
   Abstract The recent development of the Internet, especially the expanding use of social software and dynamic content generation commonly termed as Web 2.0 enables users to find information about almost every possible topic on the Web. On the downside, it becomes more and more difficult to decide which information can be trusted in. In this paper we propose the enhancement of Web 2.0 by a scalable and secure cross-platform reputation system that takes into account a user's social network. Our proposed solution *Jason* is based on standard methods of the semantic web and does not need a central entity. It enables the fast and flexible evaluation of arbitrary content on the World Wide Web. In contrast to many other reputation systems it provides mechanisms to ensure the authenticity of web content, thus, enabling the user to explicitely choose information published by trusted authors.

6. M. Hansen, "Linkage Control - Integrating the Essence of Privacy Protection into Identity Management Systems," in *Proc. of eChallenges* [Han08a].
   Abstract In the digital world, linkage of data may pose threats to the privacy of individuals. Thus, linkage control by the individuals concerned, based on transparency of the actual and planned data processing, is the main requirement to

[1] `http://www.phpBB.com`

maintain their private sphere. Today's user-centric identity management systems provide some control for users, but still lack thorough concepts of linkage control. This text introduces the phases of data processing relevant to linkage. After discussing current features of user-centric identity management concepts, an extension towards better and more comprehensive linkage control by individuals is proposed, taking into account information sources from all phases of data processing. Further, economic aspects are briefly sketched. Finally, recommendations for developers and policy makers conclude the text.

7. H. Hedbom, "A survey on transparency tools for privacy purpouses," in *Proc. of the 4th FIDIS/IFIP Summer School* [Hed08].
   Abstract This paper provides a short survey on transparency tools for privacy purposes. It defines the term transparency tools, argues why they are important and gives examples for transparency tools. A classification of transparency tools is suggested and some example tools are analyzed with the help of the classification.

7.3 Privacy of data (WP2.3)

1. C.A. Ardagna, A. Stavrou, S. Jajodia, P. Samarati, R. Martin, "A multipath approach for k-anonymity in mobile hybrid networks," in *Proc. of the International Workshop on Privacy in Location-Based Applications (PILBA 2008)* [ASJ+08].
   Abstract. The ubiquitous proliferation of mobile devices has given rise to novel user-centric applications and services. In current mobile systems, users gain access to remote *service providers* over *mobile network operators* which are assumed to be trusted and not improperly use or disclose users' information. In this paper, we remove this assumption, offering privacy protection of users' requests again the prying eyes of the network operators, which we consider to be honest but curious. Furthermore, to prevent abuse of the communication privacy we provide, we elevate traffic accountability as a primary design requirement. We build on prior work on network k-anonymity and multi-path communications to provide communications' anonymity in a mobile environment. The resulting system protects users' privacy while maintaining data integrity and accountability. To verify the effectiveness of our approach and measure its overhead, we implemented a prototype of our system using WiFi-enabled devices. Our preliminary results indicate that the overall impact on the end-to-end latency is negligible, thus ensuring applicability of our solution to protect the privacy of real-time services including video streaming and voice activated services.

2. S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, P. Samarati, "Controlled Information Sharing in Collaborative Distributed Query Processing," in *Proc. of the 28th International Conference on Distributed Computing Systems (ICDCS 2008)* [DFJ+08b].
   Abstract. We present a simple, yet powerful, approach for the specification and enforcement of authorizations regulating data release among data holders collaborating in a distributed computation, to ensure that query processing discloses only data whose release has been explicitly authorized. Data disclosure is captured by

means of profiles, associated with each data computation, that describe the information carried by the result. We also present an algorithm that, given a query plan, determines whether it can be safely executed and produces a safe execution strategy. The main advantage of our approach is its simplicity that, without impacting expressiveness, makes it nicely interoperable with current solutions for collaborative computations in distributed database systems.

7.4 Access control for the protection of user-generated data (WP2.4)

1. C.A. Ardagna, M. Cremonini, S. De Capitani di Vimercati, P. Samarati, “Managing privacy in location-based access control systems,” in *Mobile Intelligence: Mobile Computing and Computational Intelligence* [ACDS08a].
   Abstract. Preserving user data privacy is one of the hottest topics in computer security. Security incidents, faulty data management practices and unauthorized trading of users personal information have often been reported in recent years, exposing victims to ID theft and unauthorized profiling. These issues are raising the bar of privacy standards, fostering innovative research, and driving new legislations. Some approaches aimed at privacy protection deal with minimizing unnecessary release of personal information, or focus on preventing leakage of personal information while in transit or once it has been released to an authorized party. The work in this chapter addresses the latter concern in the framework of location-based services. First, we consider privacy requirements for Location-Based Access Control (LBAC) systems that require, for the provision of an online service, to evaluate conditions depending on users physical locations. We then describe obfuscation techniques that modify location information to provide user privacy protection. Finally, we illustrate a privacy-aware location-based access control system that integrates the obfuscation techniques within the access control evaluation and enforcement.

2. M. Bezzi, J.C. Pazzaglia, “The anonymity vs. utility dilemma,” in *Proc. of the Information Security Solutions Europe Conference (ISSE 2008)* [BP08].
   Abstract. The number, the type of users and their usage of the internet and phones have evolved considerably, due to the emergence of the web 2.0, the decreasing cost of portable devices, the expansion of wired and wireless internet access and the digitalization of the main entertainment media. Protecting the assets of service and software providers has been the main driver for the development of security solutions in the past ten years. However, the users/customers/citizen rights have been too often neglected since the risk related to the wrong usage of personal related information was not considered by the other stakeholders. Today, the Right to Privacy is appearing on everyone’s radar, and factors as regulations, increasing number of news stories on privacy breaches, brand damages, are forcing organizations to address user privacy as a priority. In this paper, we will briefly review the main business drivers behind the raising of privacy concerns, and outline some of the current technology solutions to address privacy requirements. Finally, we will

describe some of the future challenges in the area of privacy.

3. S. Cimato, M. Gamassi, V. Piuri, R. Sassi, F. Scotti, "A multi-biometric verification system for the privacy protection of iris templates," in *Proc. of the International Workshop on Computational Intelligence in Security for Information Systems (CISIS'08)* [CGP+08a].
   Abstract. Biometric systems have been recently developed and used for authentication or identification in several scenarios, ranging from institutional purposes (border control) to commercial applications (point of sale). Two main issues are raised when such systems are applied: reliability and privacy for users. Multi-biometric systems, i.e. systems involving more than a biometric trait, increase the security of the system, but threaten users' privacy, which are compelled to release an increased amount of sensible information. In this paper, we propose a multi-biometric system, which allows the extraction of secure identifiers and ensures that the stored information does not compromise the privacy of users' biometrics. Furthermore, we show the practicality of our approach, by describing an effective construction, based on the combination of two iris templates and we present the resulting experimental data.

4. S. Cimato, M. Gamassi, V. Piuri, R. Sassi, F. Scotti, "Privacy-aware biometrics: Design and implementation of a multimodal verification system," in *Proc of the Annual Computer Security Applications Conference (ACSAC24)* [CGP+08b].
   Abstract. A serious concern in the design and use of biometric authentication systems is the privacy protection of the information derived from human biometric traits, especially since such traits cannot be replaced. Combining cryptography and biometrics, several recent works proposed to build the protection in the biometric templates themselves. While these solutions can increase the confidence in biometric systems when biometric information is stored for verification, they have been shown difficult to apply to real biometrics. In this work we present a biometric authentication technique that exploits multiple biometric traits. It is privacy-aware as it ensures privacy protection and allows the extraction of secure identifiers by means of cryptographic primitives. We also discuss the implementation of our approach by considering, as a significant example, the combination of iris and fingerprint biometrics and present experimental results obtained from real data. The implementation shows the feasibility of the scheme in practical applications.

5. S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, G. Pelosi, P. Samarati, "Preserving confidentiality of security policies in data outsourcing," in *Proc. of the 7th ACM Workshop on Privacy in the Electronic Society (WPES2008)* [DFJ+08a].
   Abstract. Recent approaches for protecting information in data outsourcing scenarios exploit the combined use of access control and cryptography. In this context, the number of keys to be distributed and managed by users can be maintained limited by using a public catalog of tokens that allow key derivation along a hierarchy. However, the public token catalog, by expressing the key derivation relationships,

may leak information on the security policies (authorizations) enforced by the system, which the data owner may instead wish to maintain confidential.

In this paper, we present an approach to protect the privacy of the tokens published in the public catalog. Consistently with the data outsourcing scenario, our solution exploits the use of cryptography, by adding an encryption layer to the catalog. A complicating issue in this respect is that this new encryption layer should follow a derivation path that is "reversed" with respect to the key derivation. Our approach solves this problem by combining cryptography and transitive closure information. The result is an efficient solution allowing token release and traversal of the key derivation structure only to those users authorized to access the underlying resources. We also present experimental results that illustrate the behavior of our technique in large settings.

6. V. Piuri, F. Scotti, "Image processing for fingerprint biometrics via lowcost cameras and webcams," in *Proc. of the IEEE International Conference on Biometrics: Theory, Applications and Systems* [PS08b].
   Abstract. The diffusion of mobile cameras and webcams is rapidly growing. Unfortunately, images produced by these kinds of sensors during the acquisition of human fingertips are very different from the images obtained by dedicated fingerprint sensors, especially as quality is concerned. At the present stage of the research, fingerprint biometrics can be successfully achieved in real-life applications only by using dedicated sensors and scanners. In the literature a paramount quantity of methods which are extremely effective in processing fingerprints obtained by classical sensors and procedures is presented. In this paper, we investigate new techniques to suitably process the camera images of fingertips in order to produce images which are as similar as possible to the ones coming from dedicated sensors. This will allow for directly reusing the large and valuable experience presented in the literature for fingerprint recognition and verification in environments in which mobile cameras and webcams are already or can easily become available and dedicated devices are not required by the desired security level. The proposed technique encompasses a prefiltering step, the segmentation of the fingertip image, a fingertip registration phase, the dedicated processing techniques for the ridge enhancement, and a post-processing phase. In our research we tested the identification capability of the proposed methods by using a state-of-the-art, public software for minutiae extraction and matching. The effects of different registration algorithms on the identification accuracy are also discussed and the final system has been compared with the use of commercial dedicated sensors.

Bibliography

[ACDS08a] C.A. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati. Managing privacy in location-based access control systems. In L.T. Yang, A.B. Waluyo, J. Ma, L. Tan, and B. Srinivasan, editors, *Mobile Intelligence: Mobile Computing and Computational Intelligence.* John Wiley & Sons, Inc., 2008.

[ACDS08b] C.A. Ardagna, M. Cremonini, S. De Capitani di Vimercati, and P. Samarati. A privacy-aware access control system. *Journal of Computer Security (JCS)*, 16(4):369–392, 2008.

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

[AFB05] M.J. Atallah, K.B. Frikken, and M. Blanton. Dynamic and efficient key management for access hierarchies. In *Proc. of the 12th ACM CCS Conference*, Alexandria, USA, November 2005.

[AG06] A. Acquisti and R. Gross. Imagined communities: Awareness, information sharing, and privacy on the facebook. In *Proc. of 6th Workshop on Privacy Enhancing Technologies*, pages 36–58, 2006.

[AHK$^+$03] P. Ashley, S. Hada, G. Karjoth, C. Powers, and M. Schunter. Enterprise privacy authorization language (epal 1.2), 2003.

[AS01] A. Adams and M.A. Sasse. Privacy in multimedia communications: Protecting users, not just data. In *People and Computers XV - Interaction without frontiers. Joint Proceedings of HCI2001*, pages 46–69, 2001.

[ASJ$^+$08] C.A. Ardagna, A. Stavrou, S. Jajodia, P. Samarati, and R. Martin. A multi-path approach for k-anonymity in mobile hybrid networks. In *Proc. of the International Workshop on Privacy in Location-Based Applications (PILBA 2008)*, Malaga, Spain, October 2008.

[AT83] S. Akl and P. Taylor. Cryptographic solution to a problem of access control in a hierarchy. *ACM Transactions on Computer System*, 1(3):239–248, 1983.

[BBS04] D. Boneh, X. Boyen, and H. Shacham. Short group signatures. In *CRYPTO '04*, volume 3152 of LNCS, pages 45–55, 2004.

[BCC$^+$08] M. Belenkiy, J. Camenisch, M. Chase, M. Kohlweiss, A. Lysyanskaya, and H. Shacham. Delegatable anonymous credentials. Cryptology ePrint Archive, Report 2008/428, 2008. `http://eprint.iacr.org/`.

[BCKL08] M. Belenkiy, M. Chase, M. Kohlweiss, and A. Lysyanskaya. P-signatures and noninteractive anonymous credentials. In *Theory of Cryptography Conference (TCC 2008)*, Lecture Notes in Computer Science, page 18, New York,NY,USA, 2008. Springer-Verlag.

[BDF$^+$05] K. Borcea, H. Donker, E. Franz, K. Liesebach, H. Donker, and H. Wahrig. Intra-Application Partitioning of Personal Data. In *Workshop on Privacy-Enhanced Personalization*. UC Irvine Institute for Software Research (ISR), July 2005.

[Ber09] M. Bergmann. Testing privacy awareness. In *Proc. of the IFIP/FIDIS Internet Security and Privacy Summer School, Masaryk University Brno, 1-7 September 2008*. Springer Verlag, to appear/2009.

[BFG07] M.Y. Becker, C. Fournet, and A.D. Gordon. Design and semantics of a decentralized authorization language. In *20th IEEE Computer Security Foundations Symposium (CSF)*, pages 3–15, 2007.

[BN08] L. Bussard and A. Nano. Design of the 3rd version of the secse delivery platform (focused on idm). Technical Report A4.D19 v3, SeCSE Project, June 2008.

[BNN$^+$08] L. Bussard, E. Di Nitto, A. Nano, O. Nano, and G. Ripa. An approach to identity management for service centric systems. In Springer, editor, *Towards a Service-Based Internet*, volume 5377/2008 of *Lecture Notes in Computer Science*, pages 254–265, December 2008.

[BP08] M. Bezzi and J.C. Pazzaglia. The anonymity vs. utility dilemma. In *Proc. of the Information Security Solutions Europe Conference (ISSE 2008)*, Madrid, Spain, October 2008.

[BPHL$^+$06] K. Borcea-Pfitzmann, M. Hansen, K. Liesebach, A. Pfitzmann, and S. Steinbrecher. What user-controlled identity management should learn from communities. *Information Security Technical Report (ISTR)*, 11(3):119–128, August 2006.

[BPHL$^+$07a] K. Borcea-Pfitzmann, M. Hansen, K. Liesebach, A. Pfitzmann, and S. Recher. Managing one's identities in organisational and social settings. *DuD, Datenschutz und Datensicherheit*, 31(9):671–675, 2007.

[BPHL$^+$07b] K. Borcea-Pfitzmann, M. Hansen, K. Liesebach, A. Pfitzmann, and S. Steinbrecher. Managing one's identities in organisational and social settings. *Datenschutz und Datensicherheit*, 9, 2007.

[BR01] V. Buskens and W. Raub. Embedded trust: Control and learning. In Ed Lawler and Shane Thye, editors, *Group Cohesion, Trust, and Solidarity*, volume 19 of *Advances in Group Processes*, pages 167–202, 2001.

[Bra93] S. Brands. Electronic cash systems based on the representation problem in groups of prime order. In *Preproceedings of CRYPTO '93*, pages 26.1–26.15, 1993.

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

[CCS08] J. Camenisch, R. Chaabouni, and A. Shelat. Efficient protocols for set membership and range proofs. In *ASIACRYPT*, volume 5350 of *Lecture Notes in Computer Science*, pages 234–252, 2008.

[CG08] J. Camenisch and T. Groß. Efficient attributes for anonymous credentials. In *ACM Conference on Computer and Communications Security*, pages 345–356, 2008.

[CGHB08] J. Camenisch, T. Groß, and T.S. Heydt-Benjamin. Rethinking accountable privacy supporting services: extended abstract. In Elisa Bertino and Kenji Takahashi, editors, *Digital Identity Management*, pages 1–8, 2008.

[CGP$^+$08a] S. Cimato, M. Gamassi, V. Piuri, R. Sassi, and F. Scotti. A multibiometric verification system for the privacy protection of iris templates. In *Proc. of the International Workshop on Computational Intelligence in Security for Information Systems (CISIS'08)*, Genoa, Italy, October 2008.

[CGP$^+$08b] S. Cimato, M. Gamassi, V. Piuri, R. Sassi, and F. Scotti. Privacy-aware biometrics: Design and implementation of a multimodal verification system. In *Proc of the Annual Computer Security Applications Conference (ACSAC24)*, Anaheim, California, USA, December 2008.

[Cha82] D. Chaum. Blind signatures for untraceable payments. In *Advances in Cryptology – CRYPTO '82*, pages 199–203, 1982.

[Cha85] D. Chaum. Security without identification: Transaction systems to make big brother obsolete. *Communications of the ACM*, 28(10):1030–1044, October 1985.

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

[CKRS09] J. Camenisch, M. Kohlweiss, A. Rial, and C. Sheedy. Blind and anonymous identity-based encryption and authorised private searches on public key encrypted data. In *Public Key Cryptography, 12th International Workshop on Practice and Theory in Public Key Cryptosystems, PKC 2009*, Lecture Notes in Computer Science, page 19, Irvine,CA,USA, 2009. Springer-Verlag.

[CKS09] J. Camenisch, M. Kohlweiss, and C. Soriente. An accumulator based on bilinear maps and efficient revocation for anonymous credentials. In

Public Key Cryptography, 12th International Workshop on Practice and Theory in Public Key Cryptosystems, PKC 2009, Lecture Notes in Computer Science, page 20, Irvine, CA, USA, 2009. Springer-Verlag.

[CL01a] J. Camenisch and A. Lysyanskaya. Efficient non-transferable anonymous multi-show credential system with optional anonymity revocation. In Birgit Pfitzmann, editor, *EUROCRYPT 2001*, volume 2045 of *LNCS*, pages 93–118. Springer Verlag, 2001.

[CL01b] J. Camenisch and A. Lysyanskaya. An efficient system for non-transferable anonymous credentials with optional anonymity revocation. *Lecture Notes in Computer Science*, 2045, 2001.

[Cla06] S. Clauß. A framework for quantification of linkability within a privacy-enhancing identity management system. In Günter Müller, editor, *Emerging Trends in Information and Communication Security (ETRICS)*, volume 3995 of *Lecture Notes in Computer Science*, pages 191–205, Berlin Heidelberg, 2006. Springer.

[Cla07] S. Clauß. *Towards Quantification of Privacy Within a Privacy-Enhancing Identity Management System.* PhD thesis, Technische Universität Dresden, December 2007.

[CMW06] J. Crampton, K. Martin, and P. Wild. On key assignment for hierarchical access control. In *Proc. of the 19th IEEE CSFW Workshop*, Venice, Italy, July 2006.

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

[CS03] J. Camenisch and V. Shoup. Practical verifiable encryption and decryption of discrete logarithms. In Dan Boneh, editor, *Advances in Cryptology — CRYPTO 2003*, volume 2729 of *lncs*, pages 126–144, 2003.

[CS05] R. Chellappa and R. Sin. Personalization versus privacy: An empirical examination of the online consumer's dilemma. *Information Technology and Management*, 6(2-3):181–202, 2005.

[CS06] S. Clauß and S. Schiffner. Structuring anonymity metrics. In Atsuhiro Goto, editor, *Proceedings of the second ACM workshop on Digital identity management*, pages 55–62, Alexandria, Virginia, USA, November 2006. ACM.

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

[CW88] C. Camerer and K. Weigelt. Experimental tests of a sequential equilibrium reputation model. *Econometrica*, 56:1–36, 1988.

[D 7] D 7.12: Biometric behavioural profiling and transparency enhancing tools. FIDIS Deliverable. Work in progress.

[Das00] P. Dasgupta. Trust as a commodity. In Diego Gambetta, editor, *Trust: Making and Breaking Cooperative Relations*, pages 49–72. Department of Sociology, University Oxford, 2000.

[DFJ$^+$07a] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P. Samarati. A data outsourcing architecture combining cryptography and access control. In *Proc. of the 1st Computer Security Architecture Workshop*, Fairfax, VA, USA, November 2007.

[DFJ$^+$07b] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P. Samarati. Over-encryption: Management of access control evolution on outsourced data. In *Proc. of the 33rd VLDB Conference*, Vienna, Austria, September 2007.

[DFJ$^+$08a] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, G. Pelosi, and P. Samarati. Preserving confidentiality of security policies in data outsourcing. In *Proc. of the 7th ACM Workshop on Privacy in the Electronic Society (WPES2008)*, Alexandria, Virginia, USA, October 2008.

[DFJ$^+$08b] S. De Capitani di Vimercati, S. Foresti, S. Jajodia, S. Paraboschi, and P. Samarati. Controlled information sharing in collaborative distributed query processing. In *Proc. of the 28th International Conference on Distributed Computing Systems (ICDCS 2008)*, Beijing, China, June 2008.

[Dil06] B. Dillaway. A unified approach to trust, delegation, and authorization in large-scale grids. Technical report, Microsoft Corporation, 2006.

[Dir] Directive 95/46 EC. Official Journal L281,23/11/1995 pp. 31-50.

[FSSRAMM08] B. Fernandez-Saavedra, R. Sanchez-Reillo, R. Alonso-Moreno, and R. Mueller. Evaluation methodology for analyzing environment influence in biometrics. In *Proc. of the 10th International Conference on Control, Automation, Robotics and Vision (ICARCV)*, 2008.

[FWBBP06] E. Franz, H. Wahrig, A. Böttcher, and K. Borcea-Pfitzmann. Access control in a privacy-aware elearning environment. In *First International Conference on Availability, Reliability and Security*, pages 879–886, 2006.

[GAHJH05] R. Gross, A. Acquisti, and III H. John Heinz. Information revelation and privacy in online social networks. In *Proc. of the ACM Workshop on Privacy in the Electronic Society*, pages 71–80, New York, NY, USA, 2005. ACM.

[GHMP04] D.F. Galetta, R. Henry, S. McCoy, and P. Polak. Web site delays: How tolerant are users? *Journal of the Assosiaciation for Information Systems*, 5:1 – 28, 2004.

[GS08] J. Groth and A. Sahai. Efficient non-interactive proof systems for bilinear groups. In Nigel Smart, editor, *EUROCRYPT*, 2008.

[Han07] M. Hansen. Marrying transparency tools with user-controlled identity management. In *Proc. of Third International Summer School organized by IFIP WG 9.2, 9.6/11.7, 11.6 in cooperation with FIDIS Network of Excellence and HumanIT*, Karlstad, Sweden, 2007.

[Han08a] M. Hansen. Linkage Control – Integrating the Essence of Privacy Protection into Identity Management Systems. In Paul Cunningham and Miriam Cunningham, editors, *Collaboration and the Knowledge Economy: Issues, Applications, Case Studies; Proceedings of eChallenges 2008*, pages 1585 – 1592. IOS Press, Amsterdam, 2008.

[Han08b] M. Hansen. User-controlled identity management: the key to the future of privacy? *International Journal of Intellectual Property Management (IJIPM)*, 2(4):325 – 344, 2008.

[HBP05] M. Hilty, D. Basin, and A. Pretschner. On obligations. *Computer Security - ESORICS 2005*, pages 98–117, 2005.

[Hed08] H. Hedbom. A survey on transparency tools for privacy purpouses. In *Proc. of the 4th FIDIS/IFIP Summer School*. To be published by Springer, Brno, September 2008.

[HIM03] H. Hacigümüs, B. Iyer, and S. Mehrotra. Ensuring integrity of encrypted databases in database as a service model. In *Proc. of the IFIP Conference on Data and Applications Security*, Estes Park Colorado, CA, USA, August 2003.

[JPJ05] C. Jensen, C. Potts, and C. Jensen. Privacy practices of Internet users: Self-report versus observed behavior. *International Journal of Human-Computer Studies*, 63:203–227, 2005. DOI 10.1016/j.ijhcs.2005.04.019.

[KC05] P. Kumaraguru and L. Cranor. Privacy indexes: A survey of westin's studies. *ISRI Technical Report*, 2005.

[Koe06] S. Koepsell. Low latency anonymous communication - how long are users willing to wait? In *Proc. of Emerging Trends in Information and Communication Security (ETRICS 2006)*, pages 221 – 237. Springer, 2006.

[KS04] L. Kissner and D. Song. Private and threshold set intersection. Technical report CMU-CS-04-182 (2004), School of Computer Science, Carnegie Mellon University, 2004.

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

[Mac91] W.E. Mackay. Triggers and barriers to customizing software. In *CHI '91: Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 153–160, New York, NY, USA, 1991. ACM.

[MB07] M. Casassa Mont and F. Beato. On parametric obligation policies: Enabling privacy-aware information lifecycle management in enterprises. In *POLICY '07: Proceedings of the Eighth IEEE International Workshop on Policies for Distributed Systems and Networks*, pages 51–55, Washington, DC, USA, 2007. IEEE Computer Society.

[MM06] R. Mueller and U. Martini. Decision-level fusion for standardized fingerprint match-on-card. In *9th IEEE Conference on Control, Automation, Robotics and Vision (ICARCV 2006)*, 2006.

[Mos05] T. Moses. OASIS eXtensible Access Control Markup Language (XACML) Version 2.0. OASIS Standard oasis-access_control-xacml-2.0-core-spec-os, OASIS, February 2005.

[Mue01] R. Mueller. *Fingerprint Verification with Microprocessor Security Tokens*. Herbert Utz Verlag, 2001.

[NHH07] P. Norberg, D.R. Horne, and D.A. Horne. The privacy paradox: Personal information disclosure intentions versus behaviors. *Journal of Consumer Affairs*, 41:100 – 126, 2007.

[Pö9] S. Pötzsch. Privacy awareness - a means to solve the privacy paradox? In *Proceedings of the IFIP/FIDIS Internet Security and Privacy Summer School, Masaryk University Brno, 1-7 September 2008*, Lecture Notes in Computer Science. Springer Verlag, to appear/2009.

[PBPH+09] S. Pötzsch, K. Borcea-Pfitzmann, M. Hansen, K. Liesebach, A. Pfitzmann, and S. Steinbrecher. Requirements for Identity Management from the Perspective of Multilateral Interactions. In Jan Camenisch, Ronald Leenes, and Dieter Sommer, editors, *PRIME - Privacy and Identity Management for Europe*. Springer Verlag, 2009.

[PS08a] F. Pingel and S. Steinbrecher. Multilateral secure cross-community reputation systems. In S.M. Furnell S.K. Katsikas and A. Lioy, editors, *Proceedings of Trust and Privacy in Digital Business, Fifth International Conference, TrustBus*, volume 5185 of *Lecture Notes in Computer Science*, pages 69–78. Springer, 2008.

[PS08b] V. Piuri and F. Scotti. Image processing for fingerprint biometrics via low-cost cameras and webcams. In *Proc. of the IEEE International Conference on Biometrics: Theory, Applications and Systems*, Washington, DC, USA, September-October 2008.

[Rhe93] H. Rheingold. *The Virtual Community: Homesteading on the Electronic Frontier*. Perseus Books, 1993.

[RJ08] S.P. Robbins and T.A. Judge. *Organizational Behavior*. Prentice Hall, 13 edition, 2008.

[RKZF00] P. Resnick, K. Kuwabara, R. Zeckhauser, and E. Friedman. Reputation systems. *Communications of the ACM*, 43(12):45–48, 2000.

[RPM99] K. Rannenberg, A. Pfitzmann, and G. Müller. IT security and multilateral security. In *Multilateral Security in Communications*, volume 3 (Technology, Infrastructure, Economy), pages 21–29, MÃ¼nchen, 1999. Addison-Wesley.

[San87] R.S. Sandhu. On some cryptographic solutions for access control in a tree hierarchy. In *Proc. of the 1987 Fall Joint Computer Conference on Exploring Technology: Today and Tomorrow*, Dallas, USA, October 1987.

[Sch99] B. Schneier. Biometrics: uses and abuses. *Communications of ACM*, 42(8), August 1999.

[SGB01] S. Spiekermann, J. Grossklags, and B. Berendt. E-privacy in 2nd generation e-commerce: privacy preferences versus actual behavior. In *EC '01: Proceedings of the 3rd ACM conference on Electronic Commerce*, pages 38–47, New York, NY, USA, 2001. ACM.

[SGM09] S. Steinbrecher, S. Groß, and M. Meichau. Jason: A scalable reputation system for the semantic web. In *Proceedings of IFIP Sec 2009, IFIP International Information Security Conference: Security and Privacy in Dynamic Environments*, May 2009.

[Sim82] H.A. Simon. *Models of bounded rationality.* MIT Press, 1982.

[SK86] L. Sproull and S. Kiesler. Reducing social context cues: electronic mail in organizational communication. *Management Science*, 32(11):1492–1512, 1986.

[SK03] S. Steinbrecher and S. Köpsell. Modelling unlinkability. In Roger Dingledine, editor, *Workshop on Privacy Enhancing Technologies*, volume 2760 of *LNCS*, pages 32–47. Springer-Verlag, March 2003.

[SKX98] J.H. Schlichter, M. Koch, and C. Xu. Awareness - the common link between groupware and community support systems. In *Community Computing and Support Systems, Social Interaction in Networked Communities [the book is based on the Kyoto Meeting on Social Interaction and Communityware, held in Kyoto, Japan, in June 1998]*, pages 77–93, London, UK, 1998. Springer-Verlag.

[SSA06] S. Sackmann, J. Straker, and R. Accorsi. Personalization in privacy-aware highly dynamic systems. *Communications of the ACM*, 49(9), September 2006.

[Ste08] S. Steinbrecher. Enhancing multilateral security in and by reputation systems. In *to be published in Fourth FIDIS International Summer School 2008, in cooperation with IFIP WG 9.2, 9.6/11.7, 11.6; Springer 2009*, 2008.

[TC07] OASIS Web Service Secure Exchange TC. Ws-trust 1.3. Technical report, OASIS, 2007.

[Tuc65] B.W. Tuckman. Developmental sequence in small groups. *Psychological Bulletin*, 63(6):384–399, 1965.

[Var96] H.R. Varian. Economic aspects of personal privacy. *Privacy and Self-Regulation in the Information Age*, 1996.

[Wes67] A. Westin. *Privacy and Freedom.* Antheneum, 1967.

[WP00] G. Wolf and A. Pfitzmann. Properties of protection goals and their integration into a user interface. *Computer Networks*, 32(6):685–699, 2000.